Appunti sulle Reti

Routing in Internet

Ver.1 - Testo scritto da Pizzichetti Pasquale, Treviso marzo 1998, email: linopiz@iol.it

L’architettura di Internet

L’argomento End-to-End

Nelle reti con circuiti virtuali, tipo X.25 o ISDN, il controllo della trasmissione avviene salto a salto; dunque la risorsa deve essere allocata sul nodo di rete. Mentre nelle reti a datagrammi, tipo TCP/IP (stateless), il controllo avviene end-to-end; dunque l' intelligenza e' soprattutto allocata ai capi della comunicazione (terminali). Non ci sono rotte da impostare; tutti i datagrammi IP portano un indirizzo di origine e di destinazione ed ognuno viene instradato indipendentemente. Non ci sono conferme tra il terminale/client e il commutatore di rete, tantomeno alcuna garanzia che la rete non perderà pacchetti: questi devono essere confermati da capo a capo cioè “end-to-end” dal processo Transmission Control Protocol (TCP) remoto. Se la conferma non torna indietro, la stazione trasmittente rinvierà il pacchetto, che viaggerà di nuovo in tutte le vie attraverso la rete.

Il principio end-to-end descrive la divisione delle responsabilità tra la rete e suoi client: la prima fa il routing e l'altra fa il controllo.

A vantaggio delle reti a datagrammi va detto che la procedura di controllo di flusso dei dati non genera appositi pacchetti (come sui circuiti virtuali X.25), ma essa e' inserita nella conferma TCP end-to-end.

Bisogna fidarsi delle reti?

Le reti a circuiti virtuali come X.25 e quelle a datagrammi come TCP/IP hanno una complessità equivalente. Entrambe assicurano che i pacchetti vengano consegnati correttamente all’utente finale. La differenza è nell’implementazioni delle funzioni: nel terminale utente per TCP/IP e nella rete per X.25.

Dal punto di vista dell'utente: nel caso TCP/IP bisogna fidarsi di se stesso, mentre si deve fidare della rete nel caso X.25.

Oggi, sulle reti moderne, il rischio di perdere pacchetti a causa di errori di trasmissioni è molto basso. Pertanto un controllo salto a salto, tipico di X.25, sembra eccessivo: è comunque necessaria una conferma da capo a capo (end-to-end) per assicurarsi che i dati siano arrivati a destinazione e siano stati processati.

Comunque Internet include anche dei collegamenti disturbati, come vecchi circuiti telefonici o canali radio. Per questo è necessario mantenere la frequenza degli errori solo entro limiti ragionevoli: una buona regola da seguire è quella che il link peggiore non dovrebbe avere una perdita di pacchetti peggiore del 1%, altrimenti bisognerebbe controllare le apparecchiature o implementare una procedura di correzione d'errore.

Indirizzi ed interfacce TCP/IP:

Le emissioni possono essere di tre tipi:

1. *Unicast* quando sono indirizzate ad un host ben preciso sulla rete
2. *Broadcast* quando sono indirizzate a tutti gli hosts presenti sulla rete
3. *Multicast* quando sono indirizzate ad un gruppo di hosts presenti sulla rete

Gli indirizzi IP designano le interfacce di rete e non l'host.

Indirizzi speciali

0.0.0.0 = questo host su questa rete: viene utilizzato solo come indirizzo sorgente, ad esempio per effettuare il bootstrap.

255.255.255.255 = broadcast limitato, usato solo come indirizzo di destinazione indirizzato a tutti gli hosts sulla sottorete locale.

A.255.255.255 , B.B.255.255, C.C.C.255 = broadcast diretto a tutti gli hosts sulla rete A, B.B oppure C.C.C

127.0.0.1 = local host (loop locale)

224.0.0.1 = Multicast indirizzato a tutti gli hosts su questa sottorete

224.0.0.2 = Multicast indirizzato a tutti i routers su questa sottorete

224.0.0.5 = Multicast indirizzato a tutti i routers OSPF

224.0.0.6 = Multicast indirizzato a tutti i routers OSPF designati

224.0.0.9 = Multicast indirizzato a tutti i routers RIP-2

Da notare che RFC791 specifica che i routers dovrebbero sempre decrementare il valore di TTL quando questi passano pacchetti i una unità.

ICMP

Tutti i routers IP e gli hosts dovrebbero poter utilizzare questo protocollo. Il proposito di ICMP é quello di fornire riscontro sui problemi di rete.

I messaggi ICMP di tipo "*Redirect*", "*Router advertisment*" e "*Router solicitation*" vengono usati per passare informazioni di instradamento dagli/agli hosts.

Allo scopo di evitare dannosi messaggi recursivi, nessun rapporto ICMP verrà mai attivato da particolari messaggi ICMP.

Traceroute e Ping

Esistono strumenti di analisi popolari che usano IP e ICMP: *Ping* rileva la connettività end-to-end; *Traceroute* tenta di scoprire i vari ripetitori o routers lungo il cammino.

Scoprire il router locale

Quando ci sono diversi routers connessi alla rete locale, gli hosts dovrebbero selezionare normalmente quello più vicino alla destinazione. Ciò può avvenire attraverso una procedura dinamica. Infatti il protocollo ICMP prevede uno speciale messaggio, sebbene ciò si potrebbe comunque realizzare ascoltando i protocolli di routing (Rip, Ospf, ecc.).

I routers normalmente inviano notizie all' indirizzo multicast 224.0.0.1, cioé a tutti gli hosts; oppure con messaggio broadcast limitato 255.255.255.255 se il modo multicast non é supportato.

I pacchetti ICMP con le informazioni sui router vengono inviati ogni 7 minuti e queste informazioni durano mezz'ora.

Se un host é appena entrato in rete può inviare un pacchetto ICMP "Solicitation Router", all'indirizzo multicast 224.0.0.2 ovvero a tutti i router sulla sottorete; oppure con un broadcast limitato 255.255.255.255.

Gli hosts riceveranno molte indicazioni di router, ma dovranno ignorare i router non locali (non della sottorete).

Redirezione (Redirect)

La procedura di redirezione consente agli hosts di ottimizzare i cammini scegliendo il router migliore verso la destinazione sulla propria sottorete. Una volta ricevuto il messaggio di redirezione, si suppone che quell'host installi sulla propria tabella delle rotte quella voce. Ciò gli permetterà di indirizzare future comunicazioni attraverso il router giusto senza sovraccarico per la rete locale.

Buchi neri (Black Holes)

Per evitare di inviare dati ad un router "morto", in generale il principio é che un host dovrebbe ricevere qualche riscontro se quel router é ancora operativo. Per esempio attraverso la continua ricezione di conferme da una connessione TCP, oppure se il router risponde alle richieste ARP. Oppure inviando dei Ping.

DNS

Se un host ha molte interfacce, ha anche molti indirizzi. Il Dns restituirà nelle risposte tutti questi numeri IP e l'interfaccia di trasporto dovrebbe scegliere il migliore.

Protocolli di routing

RIP

Protocollo della famiglia Distanza-Vettore (distance-vector, detto di Bellman-Ford basato sull'algoritmo di calcolo Shortest Path. Incorporato inizialmente come codice di rete nello Unix della Berkley University (BSD) nel programma "Routed". Nella sua prima versione (RIP-1) era un protocollo estremamente semplice che richiedeva una minima configurazione.

Esso è stato progettato come protocollo gateway interno (IGP) per lo scambio di informazioni all'interno di un sistema autonomo (AS), cioè una rete di dimensioni relativamente limitate.

Le voci presenti nella tabella RIP sono indirizzi Internet a 32 bit e possono rappresentare hosts, reti, o sottoreti. L'indirizzo 0.0.0.0 rappresenta una rotta di default (prestabilita) verso reti fuori dal sistema autonomo (AS).

RIP usa una metrica molto semplice: la distanza è il numero dei collegamenti (Link) che devono essere utilizzati per raggiungere la destinazione, cioè il conteggio dei salti espresso da un numero intero che va da 1 a 15; il valore 16 indica infinito.

Rip supporta sia collegamenti punto a punto che reti broadcast come Ethernet.

I pacchetti RIP vengono inviati ogni 30 secondi o meno in caso di aggiornamenti provocati (Triggered update). Se la rotta non viene aggiornato entro 180 secondi, la distanza viene impostata a infinito (16) e la voce verrà successivamente rimossa dalla tabella delle rotte.

Effetti negativi di RIP: effetto rimbalzo e tempo lungo per il conteggio a infinito.

Tecniche per ridurre questi effetti:

- *Split horizon*: cioé emissione di informazioni sulla tratta su cui avvengono gli instradamento in forma semplice o con "poisoned reverse". Quest'ultima é una tecnica aggressiva che pone addirittura ad infinito la rotta su cui si ricevono gli instradamenti così da evitare il successivo conteggio ad infinito.
- *Triggered update*: tentativo di aumentare la responsabilità del protocollo richiedendo ai nodi di inviare messaggi non appena questi notano un cambiamento nella loro tabella delle rotte senza dover aspettare la fine del periodo.

Nodi silenziosi

A proposito di Routed (il demone Unix che governa l'instradamento). Esso consente agli host, così dotati, di ascoltare gli aggiornamenti dei routers ogni 30 secondi. Purtroppo con l'avvento di RIP-2, OSPF e IGRP, essi comprendono solo parzialmente questi messaggi così da costruire una tabella delle rotte incompleta.

Come al solito la cosa migliore e normale da fare é che un host deve inviare i pacchetti al router di default.

RIP-2

Novità:

- consente il routing sulle sottoreti al di fuori della rete passando informazioni sulla maschera (netmask) in parallelo con l'indirizzo. Ciò permette il supporto alla maschera (netmask) di lunghezza variabile entro la stessa rete o l'aggregazione di più indirizzi di classe C entro un indirizzo di gruppo con CIDR (Classless Internet Domain Routing).
- Autenticazione tra router pari
- Domini di routing che possono condividere lo stesso "cavo" (obsoleto)
- Indicazione del salto successivo (next-hop)
- Multicasting: esso definisce un indirizzo IP multicast di classe D 224.0.0.9 per effettuare emissioni periodiche. Non è necessario il protocollo IGMP.

Ciò da una parte solleva gli hosts sulla rete locale non interessati ai messaggi di routing emessi da RIP-2, ma dall'altra crea problemi di compatibilità con i router che usano RIP-1. Perciò è prevista (RFC 1388) una fase di transizione da RIP-1 a RIP-2 in cui dapprima RIP-1 usa pacchetti broadcast, poi sarà RIP-2 ad effettuare emissioni broadcasts ed infine RIP-2 effettuerà emissioni multicast perdendo però la compatibilità con RIP-1.

Indicazioni ulteriori per i protocolli "distance-vectors" come RIP

- E' necessario evitare la sincronizzazione involontaria delle emissioni RIP ogni 30 sec. Sulla LAN. Infatti ciò provoca congelamento nel traffico. Si può evitare ciò inizializzando i router in tempi diversi.
- E' necessario evitare emissioni di messaggi RIP con frequenza elevata su reti con circuiti virtuali come X.25 e ISDN poiché possono intasarli in caso di molte rotte. Sarebbe perfetto se i routers vicini potessere dare conferma alleemissioni RIP.
- Non è consigliabile usare RIP in una rete con diverse tecnologie di rete e quindi con molta differenza di caratteristiche (X.25, ATM). Ciò a causa della metrica usata da RIP che è troppo semplice.

Concludendo su RIP

Rip è semplice è può essere implementato in poche ore. I risultati sono accettabili se la topologia di rete è relativamente semplice e se i guasti sui collegamenti sono rari.

E' invece inadeguato per reti grandi e complesse. Esso infatti calcola nuove rotte dopo alcuni cambiamenti nella topologia di rete, ma in alcuni casi molto lentamente, contando per infinito. Durante il tempo necessario per effettuare i calcoli, la rete viene lasciata in uno stato transitorio in cui possono capitare loops che causano congestioni temporanee.

Perciò molti specialisti preferiscono usare protocolli più elaborati, come la famiglia Link-State (OSPF).

Rip è limitato, ma semplice. OSPF è all'opposto molto potente, ma piuttosto complesso.

Protocolli di routing a stato di link o Link State

Questi protocolli sono basati sull concetto di "mappa distribuita", infatti tutti i nodi hanno una copia della mappa di rete che viene regolarmente aggiornata. Questa famiglia di protocolli viene anche chiamata SPF cioè Shortest Path First.

Perché un protocollo Link-State è superiore?

1. *Veloce: convergenza senza loops*
2. *Supporto di metriche precise e se necessario multiple*
3. *Supporto di cammini multipli verso la destinazione*
4. *Rappresentazione separata di rotte esterne (al sistema autonomo)*

1) *Convergenza senza loops.*

Diversamente dall'algoritmo distance-vector, in cui il numero di passi richiesti per il calcolo di una nuova rotta è proporzionale al numero dei nodi, con l'algoritmo link-state il calcolo consiste solo in due fasi:

- una rapida trasmissione della nuova informazione attraverso il protocollo Flooding (riempimento)

- un calcolo locale

Inoltre la proprietà dell'assenza di fenomeni di loop è molto importante. Infatti subito dopo l'operazione di riempimento (flooding) e calcolo, tutte le rotte nella rete sono sane, senza loop intermedi e senza necessità di contare per l'infinito.

2) *Supporto di metriche precise e se necessario multiple.*

Possedendo una precisione di calcolo è possibile supportare più metriche in parallelo. Per esempio circa:

- *la più alta velocità*
- *il costo più basso*
- *il ritardo più basso*
- *la migliore affidabilità*

3) *Supporto di cammini multipli verso la destinazione* (Multiple path).

Nelle reti complesse, e' facile che esistano rotte quasi equivalenti verso la stessa destinazione. RIP sceglierebbe casualmente uno dei due cammini a causa dell'unica possibile voce next-hop (salto successivo) nella tabella delle rotte e quindi tutta la capacita' disponibile rimarrebbe inutilizzata.

E' stato provato che dividere il traffico su piu' cammini e' piu' efficiente; e' percio' una buona idea e dovrebbe essere implementata.

4) *Rappresentazione separata di rotte esterne.*

Grazie ad una metrica precisa, e' possibile distinguere le rotte esterne al proprio sistema autonomo (AS) attraverso dei records appropriati nel database.

<u>Il progetto di OSPF</u>

Si tratta di un protocollo Link-State con i tipici elementi:

- un database distribuito
- una procedura di riempimento o flooding
- una definizione di adiacenza
- records speciali per le rotte esterne

OSPF e' stato progettato per supportare specialmente:

1. separare hosts e routers
2. reti broadcasts, come Ethernet o FDDI
3. reti non broadcasts, come X.25 o ISDN
4. dividere reti molto larghe in piu' aree

1) Separare hosts e routers

Secondo l'algoritmo Link-State e' necessario descrivere la relazione tra ogni hosts e il router. Nel caso di una sottorete e' sufficiente indicare il numero di sottorete. Questa relazione viene chiamata: *Link to* oppure *Stub network*.

2) Reti broadcasts

Per ridurre il numero di router adiacenti sulla rete locale nel database, si designa un router "*piu' equo*". Gli altri router stabiliranno adiacenze solo con il *router designato*, sincronizzando il loro database (attraverso il protocollo Hello).

E' dimostrato che se tutti i database dei router sono sincronizzati con il router designato, allora tutti sono sincronizzati insieme e dunque non e' necessario effettuare ulteriori negoziazioni.

Cio' riduce anche i record Link-State nel database e la procedura di flooding riempimento.

Quando un router deve inviare un messaggio Link-State usa l'indirizzo multicast *224.0.0.6* ovvero destinato a *tutti i routers designati*. Se questo e' un nuovo messaggio il router designato effettuera' il flooding (allineamento) su tutte le sue interfacce indirizzando a *tutti i router OSPF* con l'indirizzo IP multicast *224.0.0.5*.

Nello stesso momento in cui viene designato il router principale viene anche designato il *Router di backup designato*. E' necessario che tutti i router mantengano adiacenze con i router designati e il suo backup. Se il router designato cade allora il backup prende immediatamente il suo posto e viene eletto un nuovo backup. Alla ricomparsa del vecchio router designato esso stabilira' nuove adiacenze con il nuovo designato e

il nuovo backup. Il protocollo Hello permette la rivelazione della caduta del router designato.

3) Reti non broadcast

OSPF applica le stesse regole applicate nelle reti broadcast

4) Aree multiple

Per evitare problemi di volume eccessivo di messaggi e tempi di calcolo elevati in reti molto grandi, si applica il principio del *routing gerarchico*. Quindi si divide in piu' parti dette *aree* attaccate ad una dorsale detta *backbone area*.

Allo scopo di tenere insieme le aree ci sono dei routers che appartenendo a piu' aree instraderanno il traffico da e per. Essi si chiamano *AREA-BORDER-ROUTER* e c'è ne sarà almeno uno in ogni area che effettua la connessione alla dorsale.

STUB-AREA e' un area dove le rotte esterne vengono riassunte da una *rotta di default*.

NOT-SO-STUBBY AREA e' un area dove tutte le rotte esterne vengono sostituite con una rotta di default *eccetto alcune*!

STUB-NETWORK descrive il collegamento tra il router e la sottorete.

Dal punto di vista del router di confine area (AREA-BORDER-ROUTER) i collegamenti esterni descrivono rotte uscenti e i collegamenti interni descrivono rotte in entrambi.

Link State Database

I router OSPF nella stessa area condividono un database composto da record Link-State. Essi rappresentano la topologia di rete e vengono utilizzate per calcolare il cammino piu' breve.

Ci sono cinque tipi di Link-state:

- *Router*
- *Network*

- *Summary for IP network*
- *Summary for border router*
- *External*

OSPF usa pacchetti IP con il campo tipo 89 ed e' composto da 3 sottoprotocolli:

- *Hello*
- *Exchange*
- *Flooding*

Bibliografia

Routing in Internet – autore Christian Huitema - casa editrice Prentice Hall

---