

Una breve descrizione dei packet-by-packet router e del metodo usato da questi per prendere decisioni sul forwarding (inoltre delle informazioni)

IP routing e forward classification

Gianluca Pedrina
Cypress Semiconductor
South Europe Staff FAE

Internet è formato da una maglia di router collegati tra loro da link. La comunicazione fra i nodi di Internet (router ed estremità-ospiti) avviene usando un protocollo di Internet normalmente conosciuto come IP. I pacchetti IP viaggiano sopra i collegamenti da un router al seguente nel loro senso di marcia verso la destinazione finale.

Ogni router (si veda la figura 1) realizza una decisione di spedizione sui pacchetti ricevuti per determinare il router del passaggio seguente.

La possibilità di spedire i pacchetti è un requisito di ogni IP router. Inoltre, un IP router può anche scegliere di realizzare una speciale elaborazione sui pacchetti ricevuti. Gli esempi di elaborazione speciale includono il filtraggio di pacchetti per motivi di sicurezza, il trasporto di pacchetti secondo requisiti di garanzia in termini di ritardi massimo prestabiliti, trattando i pacchetti a seconda della loro priorità o effettuando statistiche sul numero di pacchetti trasmessi da altri nodi della stessa rete.

Tale elaborazione speciale richiede che il router classifichi i pacchetti ricevuti in uno dei diversi flussi. Tutti i pacchetti di un flusso obbediscono a una regola predefinita e sono elaborati in un modo simile dal router. Per esempio, tutti i pacchetti con lo stesso IP Address di arrivo possono essere definiti per formare un flusso. Un flusso può anche essere defi-

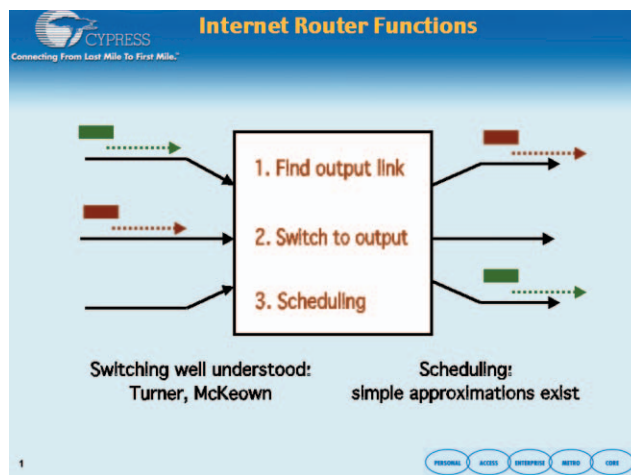


Fig. 1 - Funzioni svolte da un router Internet

nito dai valori specifici dell'IP Address della destinazione e dai valori specifici di protocollo.

Di norma, in questa sede non ci si riferirà ai router che classificano i pacchetti in base al flusso, detti router flu-aware, ma ad una categoria più ampia detta flu-unaware, ovvero router che trattano individualmente ogni pacchetto ricevuto: a questa categoria ci si riferirà come packet-by-packet router.

In particolare delle due diverse tipologie di operazioni eseguite dai router - le procedure che un IP router usa per decidere dove spedire i pacchetti (normalmente definita come route lookup o address lookup) e la procedura che un flu-aware router usa per classificare i pacchetti in flussi (parsing, routing policy, scheduling) - si tratterà brevemente solo della prima. Verrà inoltre approfondita la complessità delle procedure veloci ed effi-

cienti necessarie ai router per elaborare il maggior numero di pacchetti al secondo e, in ultima analisi, quindi aumentare la capacità e l'efficienza di Internet.

In sostanza si darà una breve descrizione dei packet-by-packet router e del metodo usato da questi per prendere decisioni sul forwarding, senza però avere ora la pretesa di descrivere come un flow-aware router classifichi i pacchetti ricevuti in flussi o spiegare gli algoritmi necessari e una metrica di valuta-

zione di questi ultimi.

Packet-by-packet Router e tabelle di routing

Un packet-by-packet router è uno speciale instradatore di pacchetti che, per definizione, deve soddisfare due funzioni basilari: essere in grado di prendere una decisione di instradamento verso il router successivo (forwarding routing) e di trasferire l'informazione dalla sua porta di ingresso alla corretta porta di uscita (switching). Quando per esempio si apre una pagina Web o si invia una e-mail, le informazioni vengono suddivise in tanti piccoli pacchetti e ogni pacchetto IP contiene al suo inizio una parte informativa (Header).

In questa parte informativa il protocollo indica se si tratta di una e-mail, di una pagina web, di un trasferimento dati ftp o un collegamento da terminale; DST

Coprocessore per la gestione del sistema di ricerca delle schede di linea

Il nuovo Vichara 81000 (CYNCP81000) di Cypress Semiconductor è un coprocessore per la supervisione delle operazioni di ricerca (Search Supervisory Coprocessor) in grado di eseguire operazioni particolarmente onerose dal punto di vista della ricerca e di gestire di ricerche multiple, consentendo quindi ai processori di pacchetti di raggiungere un throughput pari a 40 Gbps. Il coprocessore ottimizza l'ampiezza di banda del bus look-aside (LA-1) e semplifica la gestione del sistema di ricerca complessivo. Quando il coprocessore viene utilizzato insieme alla famiglia di engine per la ricerca in rete (NSE - Network Search Engine) Ayama NSE1000, è possibile raggiungere velocità dell'ordine di 266 milioni di ricerche per secondo (MSPS - Million Searches Per Second). Grazie a Vichara 81000, i processori di rete

(NPU - Network Processor Unit) sia di tipo commerciale sia implementati tramite circuiti ASIC (Application Specific Integrated Circuit) sono esonerati dall'incombenza di dover eseguire il routing (instradamento) di tabelle di ricerca e gestire funzioni legate alla policy (insieme di regole che descrivono l'azione - o le azioni - da intraprendere al verificarsi di una data condizione). In questo modo Vichara 81000 permette ai processori di pacchetti di eseguire altre funzioni richieste da una scheda di linea, con tutti i vantaggi che ciò comporta. Il coprocessore si occupa anche della gestione di funzioni di controllo, mettendo a disposizione sia le funzioni di connettività sia le risorse hardware per eseguire aggiornamenti dei controlli fuori banda delle funzioni di policing ed aging, in modo da consentire all'NPU e ai circuiti ASIC dedicati di elaborare un maggior numero di pacchetti al secondo alla medesima wirespeed (ovvero alla velocità permessa dal cavo in ingresso/uscita). Gli attuali processori di pacchetti (NPU o ASIC) eseguono funzioni quali parsing, routing, risoluzione della policy, modifica e scheduling dei pacchetti sulla scheda di linea. Oltre a ciò, il processore di controllo esegue in maniera concorrente aggiornamenti di controllo all'interno della banda sul percorso dati attraverso i processori di pacchetti. Nel momento in cui la velocità della linea aumenta da 2,5 a 40 Gbps, tali processori non sono in grado di eseguire questa

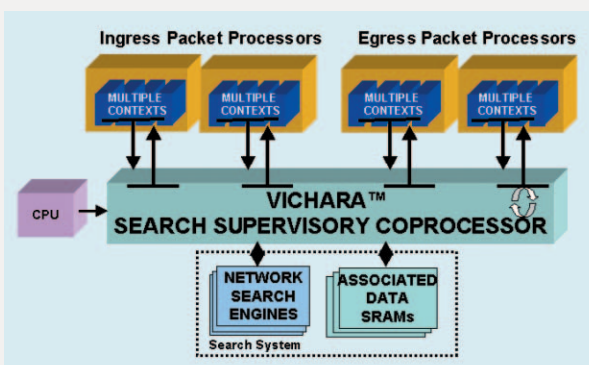
pluralità di funzioni senza dover ricorrere ad alcun ausilio esterno. Tale problema è aggravato dal fatto che vi è una crescente necessità di rafforzare le funzioni di policy per fornire servizi per i livelli più alti e rendere disponibili tabelle di ricerca più estese per gestire il crescente numero di dispositivi connessi su Internet. Un coprocessore come Vichara 81000 rappresenta un valido supporto alla soluzione di questo problema. La disponibilità di un coprocessore per la supervisione della ricerca semplifica la progettazione della scheda di linea a livello sia hardware sia software, il quanto il progettista può adattare le prestazioni del sistema di ricerca sfruttando le risorse hardware e software di Vichara. Oltre a semplificare il design della parte hardware e software dei sistemi di ricerca, il sistema di gestione della ricerca del coprocessore permette di eseguire l'operazione di ricerca all'interno di tabelle

molto ampie e profonde, in quanto è prevista la possibilità di collegare al coprocessore stesso numerosi NSE e memoria SRAM.

Alcune caratteristiche

Vichara 81000 si può a ragion veduta definire un coprocessore dei pacchetti, in quanto solleva le NPU e i circuiti ASIC dedicati dalla gestione dell'intera sequenza di operazioni di ricerca e dei dati del pacchetto e gestisce ricerche multiple nei livelli compresi tra 3 e 7. Il coprocessore opera a una frequenza di clock di 266 MHz e gestisce le operazioni di ricerca operando in congiunzione con i Network Search Engine di Cypress per gli indici di ricerca e le SRAM di tipo NoBL per i dati associati. Vichara 81000 può alternare ed eseguire più contesti in parallelo in presenza di un massimo di quattro processori di pacchetto e supporta fino a 64 contesti su ciascuna delle sue quattro porte. Per le NPU compatibili con le specifiche LA-1, tra cui quelle realizzate da Intel, Amcc e Ibm, il fatto di non doversi più occupare delle operazioni di ricerca si traduce in una maggiore efficienza del bus LA-1.

I campioni di CYNCP81000 saranno disponibili nel terzo trimestre del 2003 in package BGA a 484 pin, unitamente a una scheda di valutazione e all'interfaccia API (Application Programmino Interface) per i driver del dispositivo.



Schema a blocchi del coprocessore Vichara di Cypress Semiconductor

Addr è l'indirizzo del destinatario della informazione - nel caso si stia scaricando una pagina web da un server è l'indirizzo personale, SRC Addr è l'indirizzo di chi ha originato l'informazione (nell'esempio è il server che invia la pagina web richiesta dall'utilizzatore). Durante il lungo viaggio dal server al computer dell'utente ogni pacchetto passerà di mano in mano a molti router, ed ogni uno di questi sostituirà via via i due indirizzi di SRC Port e DST Port sino a consegnare l'informazione al browser dell'utente che la farà apparire sullo schermo.

Un caso concreto

Un esempio può servire a chiarire la complessità del problema: si supponga di essere a New York e di inviare una lettera ad un ufficio di Agrate Brianza (parallelamente può essere un server basato a New York che invia una pagina web ad un PC in un ufficio di Agrate Brianza), I due indirizzi SRD e DST saranno quello di partenza e quello di arrivo. Ora però l'utente imbuca la lettera in una cassetta per lettere, il postino la preleva e la porta al primo centro di smistamento. Comincia così il lavoro di routing: per prima cosa un impiegato (oggi anche questa operazione è automatica ma nell'esempio si eseguirà tutto il routing in modalità manuale) capirà che la lettera deve andare ad Agrate Brianza.

L'impiegato statunitense non avrà la più pallida idea di dove si trova l'ufficio di destinazione ma sarà in grado di capire che Agrate Brianza è in Europa e metterà la lettera in un contenitore per l'Europa (la prima decisione di routing è stata presa: DST Port ora è stato impostato).

La lettera prenderà un aereo e arriverà in una città europea, che potrebbe essere Londra oppure Monaco: lo smistatore successivo eseguirà la stessa operazione instradando la lettera in modo da avvicinarla sempre più a Milano e poi

infine ad Agrate Brianza dove finalmente il postino locale sarà in grado di leggere l'indirizzo finale e consegnare la lettera nella cassetta postale appropriata. L'esempio può sembrare banale ma è chiaro che ogni smistatore (router) è solo in grado, in base all'indirizzo finale DST Addr, di scegliere la migliore destinazione intermedia per avvicinare la lettera alla posta di smistamento finale. Tornando all'esempio della pagina web, il pacchetto di informazioni passerà di router in router e ognuno di questi, in base alla sua tabella di forwarding, lo instraderà al router successivo sino ad arrivare al router del provider

Finora è stato stabilito che ogni operatore, postino o smistatore deve svolgere un lavoro ben definito: trovare l'indirizzo dello smistatore o postino successivo nella catena e spostare la lettera nella cassetta di ingresso del successivo. Banalmente, dal piccolo router che si trova nell'ufficio dell'utente al router di grandi dimensioni che smista le informazioni tra Europa e Stati Uniti la decisione da prendere è la stessa. La risposta dei costruttori di router non poteva che essere scalabile: un piccolo router avrà al suo interno un microprocessore che leggerà e sostituirà l'indirizzo DST Port con il più simile al finale DST

Addr; il mastodontico router tra i due continenti avrà una batteria di processori specializzati che si avvarranno anche di dispositivi speciali esterni di ricerca e sostituzione degli indirizzi in grado di arrivare ogni uno a 266 milioni di ricerche al secondo (si veda l'apposito riquadro) coadiuvati da altri chip capaci di coordinare questi NSE (Network Search Engine) e permettere loro di lavorare in parallelo per arrivare a migliaia di milioni di ricerche al secondo. In conclusione, si è cercato di dare una semplice spiegazione alla richiesta di

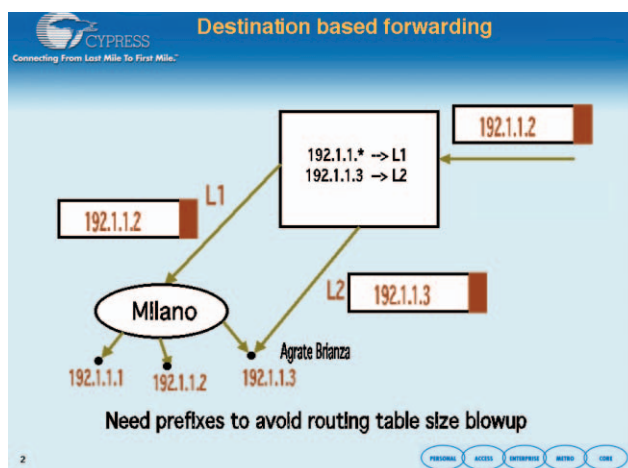


Fig. 2 - Esempio di forwarding basato sulla destinazione

appropriato oppure al router finale dell'utente. Se l'indirizzo dell'utente è per esempio 193.11.12.13, il pacchetto, in partenza da una rete 157.23.24.25 risalirà la sua catena di routers sino ad arrivare al router capace di instradarlo nella mia direzione.

È ovvio comprendere a questo punto quanto questa capacità dei router normalmente definita come route lookup o address lookup sia importante e, considerando la velocità di crescita del numero di utilizzatori di Internet, renda il compito di queste macchine estremamente complesso.

I nodi della rete aumentano, le tabelle di routing crescono e la velocità di esecuzione non deve diminuire: anzi, ci si aspetta di avere accessi sempre più veloci e maggiori performance della rete.

spiegare che cosa è un router, ma, come accennato all'inizio dell'articolo, il problema cresce con le aspettative. I nuovi router oggi prendono anche decisioni più complesse legate alla richiesta di diversificare i servizi offerti come maggiore priorità in alcuni casi o protezione dei dati trasmessi su reti speciali. Tutto questo non si discosta dall'esempio di base che resterà valido anche al crescere della complessità della rete e dei servizi offerti ma deve solo fare riflettere la crescita della complessità di questi sistemi nei prossimi anni anche considerando il tasso di crescita delle installazioni broadband (internet veloce) in ambiente domestico.

Cypress Semiconductor
Reader Service n°10