

Un primo bilancio sulla cache www CILEA

P. Tentoni

CILEA, Segrate

Abstract

In un articolo sullo scorso numero del Bollettino, si dava l'annuncio dell'adesione del CILEA al progetto per la realizzazione di un sistema di cache nazionale. A distanza di qualche mese dalla creazione del server www cache CILEA tentiamo un'analisi, anche se parziale, del successo di questa iniziativa.

La situazione nazionale

Nella scorsa riunione del gruppo a livello GARR per il coordinamento delle iniziative di caching www, è emersa la necessità di *valutare in maniera significativa il successo* che il progetto, seppure ancora non pienamente sviluppato, sta producendo.

La sensazione diffusa, a livello dei gestori di tutti i sistemi di caching, è infatti che vi sia uno spontaneo interesse da parte del pubblico (i client) riguardo l'utilizzo di un proxy con funzione di cache, interesse evidentemente in qualche modo catalizzato dalle performance complessivamente buone di tale sistema, almeno rispetto ai tempi di risposta cui la rete ci ha abituato.

Il fatto di ritrovare più velocemente una buona parte degli URL richiesti (siano essi pagine HTML, FTP o Gopher), ha indotto molti a modificare spontaneamente la definizione del proprio browser in modo da utilizzare il server cache più vicino (o il più noto in quel momento).

Del resto si sa che le voci sulla rete si spargono a macchia d'olio, arrivando prima ancora che venga fatta pubblicità ufficiale ad un servizio.

E' così che molti gestori di cache si sono visti arrivare client sempre più numerosi, richiamati come api sul miele dalle performance incoraggianti delle cache. Questo è sicuramente già di per sé un ottimo indice del successo dell'iniziativa.

Anche per la cache del CILEA si è verificato questo fenomeno. In assenza quasi totale di pubblicità al neonato servizio, che era ancora in fase di test, comparivano ogni giorno client nuovi (sia di tipo browser, sia di tipo cache-server figli), senza che questi venissero in alcun

modo né cercati, né sollecitati ad aderire all'iniziativa, né informati direttamente da alcun gestore. Il servizio è stato scoperto dagli utenti ed il loro numero si è innalzato progressivamente fino ad arrivare oggi ai 145 client di tipo browser (o non ICP) e a 10 client ICP (ovvero cache figlie o sorelle).

Questo ha portato ad un rapido e progressivo arricchimento della cache ed ad una sua resa che si aggira oggi quotidianamente su un **HIT rate del 30-35%** in termini di oggetti richiesti, ed un Hit rate in termini di Mbyte trasmessi dell'ordine del 25% rispetto al volume totale del traffico transitato dal server che mediamente è sui 700-800 Mbyte al giorno, per la sola cache del CILEA. Di questo traffico la porzione esclusivamente nazionale (cioè in .it) corrisponde, sempre per la cache CILEA, al 20-25% del totale del traffico passante.

The screenshot shows a Netscape browser window displaying the Squid Statistics page for wwwcache.cilea.it:8080. The page contains two tables summarizing request method and protocol usage.

Summary of Request Method Usage

Method	UDP COUNTS			TCP COUNTS			TCP BYTES		
	counts	%all	%hit	counts	%all	%hit	Mbytes	%all	%hit
ICP_QUERY	221676	100%	20%	0	-	-	0.00	-	-
GET	0	-	-	78746	100%	37%	806.56	100%	23%
POST	0	-	-	288	0%	-	1.32	0%	-
HEAD	0	-	-	6	0%	-	0.00	0%	-

Summary of Protocol Usage

Protocol	UDP COUNTS			TCP COUNTS			TCP BYTES		
	counts	%all	%hit	counts	%all	%hit	Mbytes	%all	%hit
http	218392	99%	20%	78170	99%	37%	565.75	70%	29%
ftp	3153	1%	3%	723	1%	12%	241.82	30%	8%
gopher	131	0%	1%	31	0%	10%	0.06	0%	12%
cache_object	0	-	-	116	0%	-	0.26	0%	-

Non ci vuol molto a capire che il risparmio di banda internazionale derivante dall'azione della sola cache del CILEA è, supponendo, come è in effetti, che la quasi totalità delle richieste avvenga durante le 12 ore dalle 8 alle 20, così calcolabile:

800Mbyte	traffico complessivo
25% (200Mbyte)	traffico solo Italiano
600Mbyte	traffico internazionale

25% HIT sui Mbyte equivalgono a **150 Mbyte/12ore.**

Quindi si ha un throughput di 30 kbps di solo traffico internazionale, concentrato sulle ore di massimo intasamento.

A prima vista potrebbe parere un numero insignificante, ma questa rappresenta la resa della sola macchina del CILEA in questa fase in cui esiste ancora a nostro avviso spazio per la crescita.

Se si ipotizza una resa media analoga anche degli altri quattro root server nazionali (alcuni avendola superiore altri inferiore, ma tutti non al massimo delle loro capacità), si arriva ad un throughput di 150 Kbps, che è, per esemplificare, pari al 9% della capacità reale di una connessione 2 Mbps.

Inoltre il conteggio della resa dei sistemi di caching non dovrebbe essere limitato ai soli HIT dei root server, che proprio per la loro natura di root server, per quanto grossi possano essere, hanno tuttavia una bassa resa rispetto agli HIT di un server di ultimo livello.

E' questo ad esempio il server dell'ente o università che deve soddisfare le esigenze dei propri client e che si è posto nella gerarchia nazionale come "figlio" dei root cache.

Nel caso di server periferici infatti si assiste di solito ad un incremento delle percentuali di HIT, che *arrivano facilmente oltre il 40%*. Tutto questo traffico resta confinato ancor più vicino al client, non arrivando neppure al successivo livello della gerarchia, ed è pertanto un ulteriore risparmio di cui si deve tenere conto.

In particolare la rete dei figli del wwwcache CILEA comprende (per ora, visto che la situazione è comunque dinamica) alcuni server di una certa rilevanza. Eccone i nomi insieme ad una indicazione del traffico che essi generano (dati del 10 aprile 1997) verso il CILEA e che è solo una parte (quella dei MISS) di quello realmente trattato dalla cache periferica:

it.uni-bocconi.sm.dns	191.88
it.polimi.cdc.master	176.01
it.cefriel.rap3	133.50
it.enel.cra.willy	24.31
it.hsr.www-n	24.18

Per chi volesse prendere visione dei dati di traffico relativi alle cache nazionali è sufficiente consultare l'URL:

<http://www.cache.garr.net/#RootCacheInside>

e quindi selezionare il nome del root cache che interessa.

