

Manutenzione e potenziamento del Servizio FTP-Mirror e Web-Cache Nazionale GARR

Paola Tentoni

CILEA, Segrate

Abstract

Il potenziamento hardware dei sistemi di mirroring è una delle attività previste dal progetto nazionale di FTP-Mirror e Web-Cache in ambito GARR per il 2003. I mirror nazionali, infatti, hanno ottenuto un gradimento considerevole e crescente da parte delle utenze GARR nel corso di questi due anni, come testimoniano i dati relativi al loro utilizzo, contribuendo, grazie all'avvicinamento dei contenuti, ad un buon risparmio di banda.

Keywords: GARR, Cache, Mirror.

Il progetto Web-CACHE [1] ed FTP-mirror [2] GARR nasceva negli anni 2001-2002 per diffondere il ricorso al caching web ai livelli periferici, ma anche per fornire contenuti, sui siti di Mirror nazionali, di interesse generale per la comunità di utenti, ed in qualche modo aperti alle richieste provenienti dagli stessi utilizzatori finali. Le ragioni di tale investimento stavano innanzitutto nell'obiettivo del risparmio di banda sulle tratte internazionali e nazionali per tutto il traffico potenzialmente ripetitivo indotto sulla rete dagli utenti stessi. Effetto secondario era poi l'avvicinamento dei contenuti a questi ultimi (tempi di accesso più bassi, nessuna congestione sui server). L'iniziativa nazionale svolgeva pertanto da guida, comportando una razionalizzazione del sistema di caching web GARR, ma dava soprattutto la possibilità di ampliare le capacità locali dei server dipartimentali o d'Ateneo mettendo a disposizione un livello centrale di caching. Tale sistema era aperto al peering, e collocato sul backbone, in prossimità dei nodi utente afferenti ai POP, dei link internazionali e dei punti di contatto con gli ISP commerciali italiani.

E' chiaro comunque che, nonostante l'ampliamento della banda complessiva disponibile oggi sul backbone e sui link internazionali, continuano a sussistere le ragioni, anche se diverse, per le quali non solo ha senso, ma è anche necessario ricorrere oggi ad una ben strutturata e localizzata gerarchia di

cache in ambito GARR. Parallelamente all'ampliamento dei backbone e dei link internazionali si ha infatti un potenziamento continuo delle capacità di banda degli end user (sia da casa, con connessioni sempre più veloci, sia soprattutto dagli uffici, cablati in Gigabit Ethernet o FastEthernet anche sul desktop), nonché della capacità di throughput delle singole macchine. Dunque è ancora sentito l'effetto delle cache nel senso di una riduzione dei tempi di risposta nella consultazione di pagine Web remote, e della riduzione del traffico indotto dalle richieste ripetitive dirette agli stessi siti ed alle stesse pagine, soprattutto essendo gli interessi largamente comuni, in una rete scientifica di scala nazionale.

A questo motivo, ritenuto ormai poco importante in situazioni di non congestione dei link, si è recentemente aggiunto quello più moderno del *content filtering*, che le cache possono svolgere, soprattutto se collocate vicino all'utente finale. Mediante il content filtering si possono evitare, con un intervento rapidissimo dei gestori, pagine pericolose (contenenti virus, come ad esempio il famoso nimda o red-code), la cui diffusione velocissima rischia, se non si dispongono di mezzi a livello centrale, di mettere in ginocchio reti intere.

Va menzionato inoltre il lavoro svolto all'indomani dell'11 settembre 2001, in cooperazione con il NOC, [3] per ristabilire il traffico web ed ftp interrotto bruscamente dopo il crollo del World Trade Center, mediante le

cache nazionali verso l'unico link attivo sugli USA. Le cache nazionali e dipartimentali furono allora il tramite per un *rerouting temporaneo*, ma veloce da attuare, verso il link disponibile subito, durante l'emergenza.

Oltre al servizio di caching nazionale GARR è nato, all'ombra del progetto di caching, quello di realizzazione di un sistema efficiente e centrale di FTP mirroring, sentito in modo particolare dai gestori di centri di calcolo e da chi deve continuamente aggiornare prodotti e macchine, rincorrendo l'ultima patch per la sicurezza, introvabile o irraggiungibile in caso di urgente bisogno. Oggi infatti, quando giungono le segnalazioni per i problemi di sicurezza, queste non hanno più una caratteristica tale da consentire al gestore il tempo necessario di procurarsi con la dovuta calma la patch e di installarla. Gli ultimi problemi di sicurezza sono stati infatti tali da generare delle vere e proprie corse alla patch per risolverli, pena l'obbligo di disconnessione del server incriminato dalla rete o il non funzionamento dell'intera rete (ad es. per il worm Sql-Slammer). Dunque l'aver "vicino" e disponibile un mirror ad elevate capacità di throughput è divenuto tanto

strategico quanto quello di un CERT [4] efficiente nel segnalare il problema. A tale proposito sarebbe auspicabile un più stretto coordinamento con il CERT stesso.

Si aggiunge a questo l'obiettivo di ottenere, anche mediante l'utilizzo dei mirror, un considerevole risparmio di banda. Le statistiche attuali (fig.1) già confermano ampiamente questo dato [5] e ci spingono, proprio per questo, a cercare di estendere i contenuti disponibili al massimo della capacità dei nuovi hardware, tenendo conto delle richieste degli utenti finali. Si prevede quindi per l'anno 2003, entro la fine di marzo, la sostituzione degli attuali server con 4 nuovi server (server X-series 345 IBM, con un processore Intel Xeon 2.4GHz, espandibile a due, doppia interfaccia 10/100/1000Mbps, 1 Gbyte RAM, 4x73 Gbyte di dischi a 10.000rpm, sistema operativo Linux-Debian), dedicati al mirroring, e con funzione di DNS locale, a supporto dei server di caching. La sostituzione delle macchine, installate e configurate con i nuovi contenuti sin da fine 2002, è già in corso (Milano e Bologna sono attive nel momento in cui stiamo scrivendo) e verrà completata entro il primo quarto del 2003

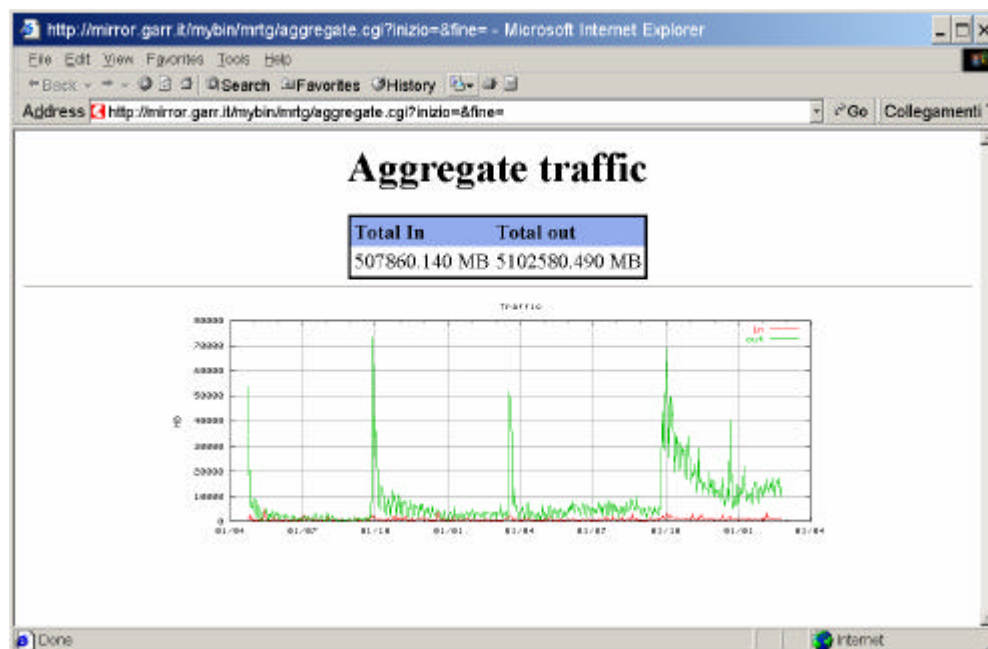


Fig. 1 –Statistiche aggregate

I precedenti sistemi, infatti, avevano ormai raggiunto la massima capacità di espansione, a meno di sostituire tutti i dischi con altri di capacità superiore. Ma l'elevata instabilità degli hardware ha suggerito la loro sostituzione piuttosto che un ulteriore investimento su di essi. L'interfaccia Gigabit Ethernet potrebbe portare ad un ulteriore aumento della banda utilizzabile. Inoltre la disponibilità del supporto di IPv6 nel software prescelto rende possibile un loro utilizzo anche nell'ambito della sperimentazione del nuovo protocollo sulla rete GARR.

Relativamente al sistema di caching invece, non ritenendo per ora giustificato un upgrade hardware ulteriore, ci si concentrerà nell'impegno di diffusione della conoscenza sul caching, soprattutto in relazione alle problematiche del content filter, alla gestione delle access-list ed al trattamento delle esclusioni necessarie per siti particolari, quali quelli delle riviste elettroniche che basano l'autorizzazione all'accesso sugli indirizzi IP del richiedente. Spesso, infatti, queste difficoltà, certamente non insormontabili, inducono gli utenti finali ad escludere l'utilizzo delle cache anche locali, ed i gestori di cache dipartimentali ad evitare la gerarchia nazionale.

A questo scopo è nostra intenzione studiare la realizzazione di un corso on-line per gestori di cache, tutto basato su strumenti OpenSource, per l'apprendimento di nozioni base relative la gestione dei sistemi di caching Squid, pensati all'interno della nostra gerarchia nazionale.

Bibliografia

- [1] URL: <http://cache.garr.it> - sito di riferimento per il caching nazionale GARR
- [2] URL: <http://mirror.garr.it> - sito di riferimento per il servizio FTP mirror nazionale GARR
- [3] URL: <http://www.noc.garr.it> - Network Operating Center GARR
- [4] URL: <http://www.cert.garr.it> - CERT GARR
- [5] URL: <http://mirror.garr.it/mybin/mrtg/lista.cgi> - Statistiche cumulative di traffico mirror.