Sulla risoluzione degli indirizzi IP
Parte III – Domain Name System

Stefano Bonacina (*), Francesca Giuratrabocchetti (**), Davide Stefanoni (***)

(*) Dottorato di Ricerca in Bioingegneria – XVI ciclo – Politecnico di Milano
(**) CILEA, Segrate
(***) già Politecnico di Milano e Medical Informatics Training Program of the National Library of Medicine (NLM) at the National Institute of Health (NIH), Bethesda, MD, USA

Abstract
Con questo articolo continua la pubblicazione a puntate della monografia sul riconoscimento degli indirizzi IP. Il lavoro è originato dalla attività di dottorato di Stefano Bonacina, dottorando in Bioingegneria del Politecnico di Milano. Il documento originale è stato curato e rivisto nel suo complesso dagli specialisti di reti del CILEA, anche per tenere conto delle attuali evoluzioni delle soluzioni software.
Questa terza parte riprende e conclude la descrizione del Domain Name System, il servizio che, sulla rete Internet, si occupa della traduzione degli indirizzi IP in nomi a dominio e viceversa. Le prime due puntate sono state pubblicate sui numeri 85 e 86 del Bollettino del CILEA.

Keywords: Telematica, TCP/IP, DNS, Reti.

Metodi di interrogazione del server Domain Name System

Ciascuna zona o gestore di un dominio mantiene un proprio database di informazioni attraverso un server Domain Name System (DNS) primario, che può essere interrogato da qualsiasi client sulla rete. I client e il name server comunicano tra di loro con l'applicativo DNS. Le applicazioni client accedono alle informazioni del DNS tramite un programma chiamato *name resolver*.
Data la brevità e la semplicità delle richieste e delle relative risposte, si utilizza il protocollo User Datagram Protocol (UDP). Il protocollo Transmission Control Protocol (TCP) è usato raramente per questo scopo. La porta impiegata dal servizio DNS è la numero 53.

Le specifiche sono contenute nel **RFC1034**, le specifiche dell'implementazione nel **RFC1035**. Concettualmente, la risoluzione di un nome di dominio segue dall'alto verso il basso, partendo dal server radice e procedendo verso i server collocati nelle foglie. Ma questo sistema porterebbe all'inefficienza per evidenti motivi, primo tra i quali l'intasamento del server radice a cui tutti andrebbero a fare riferimento. Le interrogazioni partono quindi dai server locali per poi risalire di gerarchia nel caso non si riesca a rispondere all'informazione cercata.

In qualunque sistema avente un database distribuito un determinato name server può essere palesato da una query rivolta a qualche altro name server. I due approcci generali per trattare questo problema, di interrogazione al DNS, sono:

- il **metodo ricorsivo**, in cui il server interpellato per primo rimanda l'interrogazione o query formulata dal client ad un altro server; l'intero compito è lasciato al DNS, anche il fornire la risposta. Quando un server riceve una query per la quale non ha informazioni, deve autonomamente contattare un altro *name server*, in grado di fornire l'informazione richiesta.
- il **metodo iterativo,** in cui il server indirizza il client ad un altro server e lascia al client stesso il sottoporre la query ad un altro server. Ogni volta che si interroga in questo modo un server DNS, nel caso

questo non abbia l'informazione cercata, viene restituito l'indirizzo del prossimo server a cui inoltrare la richiesta.

Un server DNS è in grado di gestire i riferimenti, per cui può usare entrambi i metodi per rispondere ad una richiesta (Figura 1).

Il resolver usa invece il metodo ricorsivo perché non è in grado di seguire i riferimenti, quindi affida l'intera operazione al suo DNS. I server radice sono non ricorsivi: se non possiedono la risposta, forniscono gli indirizzi di chi la possiede. I server DNS dei provider sono ricorsivi nei confronti dei computer dei propri utenti perché cercano la risposta e quando la trovano la restituiscono al client. I root name server contengono gli indirizzi IP di tutti i name server autoritativi (primario e secondari) per tutti i domini di primo livello.

Quindi alcune query possono risultare in una query al root name server, seguita da una query al server il cui indirizzo è stato fornito dal root name server.

I pacchetti di richiesta e di risposta, Figura 2, del DNS contengono i seguenti campi [1]:

- **Header**: Indica se il messaggio è una richiesta o una risposta, contiene alcuni flag, i codici di errore ed altre informazioni relative al pacchetto.
- **Question**: contiene la domanda per il DNS. Ogni query ad un DNS è formata da tre parti: il nome da ricercare, il tipo della query e la classe della query. I tipi più comuni di query corrispondono ai vari tipi di Resource Record (RR): A, NS, CNAME, PTR [1]. La classe è sempre IN (internet).
- **Answer**: contiene un elenco di RR che rispondono alla domanda.
- **Authority**: contiene un elenco di RR NS di server DNS che portano più vicino alla risposta.
- **Additional**: contiene un elenco di RR con informazioni utili per rispondere alla domanda, anche se non si tratta della risposta.

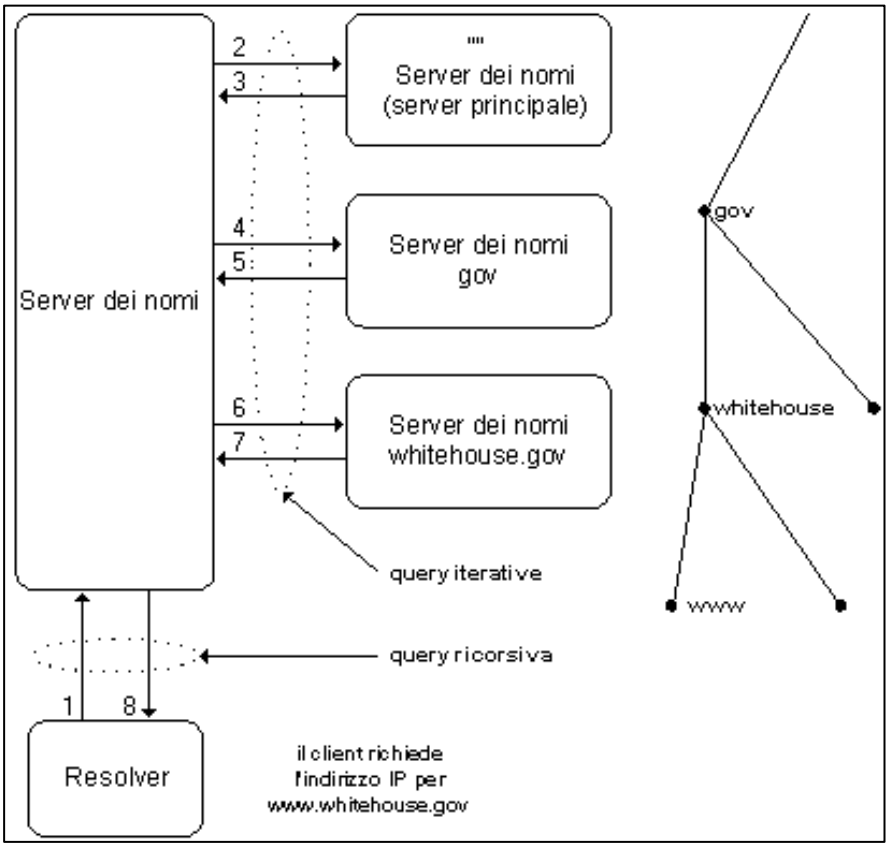

Figura 1 - Procedimento di interrogazione ricorsivo ed iterativo

Header	*PCODE=SQUERY, RESPONSE, AA*
Question	*QNAME=SRI-NIC.ARPA., QCLASS=IN, QTYPE=A*
Answer	*<EMPTY>*
Authority	*<EMPTY>*
Additional	*<EMPTY>*

Figura 2 - Schema di un pacchetto di risposta

Tra i campi dell'header si trova il flag Authoritative Answer (AA). Quando è attivo, nei messaggi di risposta, indica che il DNS che ha generato la risposta (direttamente o indirettamente) è il server autoritativo per quella zona. Se la risposta viene memorizzata nella cache del DNS locale, effettuando una seconda volta la stessa richiesta, si otterrebbe la stessa risposta ma con il flag AA non attivo, ad indicare che la risposta non è autoritativa. Una particolarità dei messaggi di risposta è che il campo Authority, contenente i RR che portano verso i server autoritativi, viene riempito anche se il messaggio è una risposta non autoritativa. In questo modo vengono comunicati quali sono i server DNS a cui rivolgersi per avere la risposta autoritativa.

Risoluzione inversa: dagli indirizzi IP ai nomi dei calcolatori

Il procedimento inverso che effettua la ricerca del nome del dominio a partire dall'indirizzo IP è solitamente chiamato *risoluzione DNS inversa* (*reverse DNS lookup*). Ottenere il nome di una macchina a partire dall'indirizzo IP è un'operazione complessa, ma serve per produrre un output maggiormente comprensibile nei file di log, o per effettuare controlli di autenticazione.

Per ricavare un indirizzo IP, dato il nome mnemonico, si può seguire la struttura gerarchica dell'albero dei domini nel modo visto in precedenza. Per effettuare, invece, l'operazione inversa non è più possibile seguire questa gerarchia. Per rendere possibile questo servizio è stato riservato il dominio speciale "in-addr.arpa", chiamato anche dominio inverso. La desinenza "arpa" è un retaggio del passato: Internet era originariamente denominata ARPAnet. I nomi dei domini sono organizzati in modo tale da avere la parte più significativa a destra, mentre gli indirizzi IP, nel formato decimale, hanno i byte più significativi a sinistra, è necessario creare il nome di dominio inverso mettendo i numeri dell'indirizzo IP in ordine inverso e aggiungendo in coda la stringa "in-addr.arpa". In questo modo viene rispettata la natura gerarchica del DNS. Un indirizzo IP in notazione decimale ha il seguente formato:

aaa.bbb.ccc.ddd

Per formare il nome di un dominio speciale in-addr.arpa occorre riscrivere l'indirizzo nella forma:

ddd.ccc.bbb.aaa.in-addr.arpa

Un esempio di dominio in-addr.arpa può essere 34.16.1.151.in-addr.arpa. Quando viene effettuata una richiesta di traduzione da IP a nome, viene effettuata una normale ricerca del nome di dominio. Ad esempio se in Unix si digita il comando "nslookup 151.1.16.34" viene eseguita una ricerca per il nome di dominio "34.16.1.151.in-addr.arpa". Per poter funzionare sono presenti nei DNS i RR di tipo PTR il cui funzionamento ricalca a grandi linee quello dei record A. Non per tutti gli indirizzi IP è possibile trovare i corrispondenti nomi a dominio, ne troverà solamente uno, non esiste infatti una corrispondenza biunivoca. Ad ogni domain name corrisponde un solo indirizzo IP, ma non viceversa.

Algoritmi di "caching"

I server DNS utilizzano solitamente algoritmi di caching, memorizzazione tampone, per ottimizzare i tempi della ricerca. Ogni server mantiene in una memoria cache i nomi recentemente usati e il modo in cui ha ottenuto queste informazioni. Quando un client richiede al DNS di risolvere un nome di dominio, il server prima di rivolgersi ai DNS di livello superiore controlla

nella sua memoria cache se quel nome è stato risolto recentemente. In caso affermativo le informazioni contenute nella cache vengono riportate al client marcandole come non autoritative e fornendo in aggiunta anche il nome del server da cui erano state ottenute. Il client riceve in questo modo una risposta molto velocemente, ma l'informazione potrebbe anche essere non più valida. Se l'efficienza è importante per il client questo accetterà la risposta non autoritativa. Nel caso invece sia importante l'accuratezza potrà scegliere di controllarne la veridicità presso l'autorità appropriata. La memorizzazione nella cache migliora notevolmente i tempi di risposta degli utenti, in quanto la maggioranza delle richieste al DNS è rappresentata da query ripetute.
Si possono distinguere principalmente due tipi di server DNS: autoritativo e di sola cache.
Un **server DNS autoritativo** contiene informazioni di database proprie e mantiene copie indipendenti delle informazioni sul dominio - i nomi e gli indirizzi IP dei server locali e di altre periferiche - nel proprio database interno. Si può gestire questo database per l'uso privato dei propri utenti (nel caso in cui non si abbiano server pubblici su Internet), oppure lo si può rendere visibile al resto della rete Internet. Per essere autoritativo e visibile a tutta la rete Internet, il server DNS deve essere connesso a Internet a tempo pieno in modo che possa essere interrogato da tutto il mondo esterno.
Un **server DNS di tipo "cache-only"** è un server DNS autoritativo con funzioni limitate: si usa lo stesso software per entrambi i tipi di server, ma l'assenza di un database del dominio obbliga il software del server ad operare in modalità cache.

La registrazione di un nome a dominio

Ogni oggetto connesso a Internet è identificabile attraverso un indirizzo IP e, anche se non sempre, un nome. Analogamente all'Allocation Registry e al Routing Registry è stato anche creato il **Domain Registry Database**, che contiene le informazioni relative ai nomi dei domini IP.
I Top Level Domain, gestiti dalle **Registry Authority** sono quelli generici, indicati nelle Tabelle 1 e 2, e quelli geografici (country code), per esempio ".it" per l'Italia, ".ch" per la Svizzera, ".de" per la Germania.

Sigla del TLD	Descrizione
.ARPA	Vecchio stile Arpanet
.COM	Organizzazioni Commerciali USA
.EDU	Università USA
.GOV	Organizzazioni Governative USA
.INT	Organizzazioni Internazionali
.MIL	Enti Militari USA
.NATO	Ambito Nato
.NET	Network
.ORG	Organizzazioni No-Profit

Tabella 1 - Sigle dei Top Level Domain e loro significato

Sigla del gTLD	Descrizione
.BIZ	Per uso commerciale di aziende e individui.
.INFO	Per uso senza restrizioni, generale.
.NAME	Per uso limitato alla registrazione di individui.
.PRO	Per uso limitato a professionisti (commercialisti, avvocati, medici) ed associazioni di professionisti.
.AERO	Per uso riservato alle compagnie aeree.
.COOP	Per uso riservato a cooperative aziendali.
.MUSEUM	Per uso riservato ai musei.

Tabella 2 - Sigle dei sette Top Level Domain, disponibili dal 2001, e loro significato

Le Registry Authority hanno il compito di mantenere documentate all'interno dei Registry tutti i propri sottodomini. Attualmente la maggior parte delle Registry Authority garantisce la documentazione dei propri sottodomini fino al livello 3; d'altra parte, sono pochi gli utenti che chiedono all'Authority la registrazione di domini a livelli inferiori.
A differenza dell'Allocation Registry Database, per il quale IANA ha un ruolo di coordinamento mondiale, non esiste un coordinamento gerarchico formalizzato per il Domain Registry Database e per il Routing Registry Database. In particolare, le organizzazioni *proprietarie* dei TLD devono mantenere un loro Domain Registry.
In Europa, RIPE gestisce il database delle reti IP europee, dei domini, delle informazioni di routing e delle persone di riferimento, in realtà da qualche anno c'è stato un passaggio delle informazioni sulle persone dal database del RIPE a quello della Registration Authority. Oggi si richiede un NIC-HANDLE (codice che identifica

univocamente una persona, costituito di norma dalle iniziali del nome e del cognome, da un numero progressivo e da una sigla identificativa dell'ente che rilascia che codice) direttamente alla Registration Authority e non più al RIPE. Tale database e' noto come RIPE Network Management Database o anche semplicemente come *RIPE Database*.

Come registrare un nome a dominio in Italia

La Registration Authority (RA) [2] è responsabile dell'assegnazione dei nomi a dominio e della gestione dei registri e del nameserver primario per il Top Level Domain ".it" , country code "IT" (ISO 3166).
I servizi forniti dalla RA sono riservati ai provider/maintainer, cioè a quelle organizzazioni che intendono registrare domini per conto terzi, o a quelle persone fisiche o giuridiche che intendono gestire direttamente le proprie reti, senza passare da un provider/maintainer.
Per diventare provider/maintainer è necessario stipulare un contratto con la Registration Authority Italiana.
La RA Italiana ha inoltre il compito di gestire i registri operativi del Top Level Domain ".it". Le modalità operative generali e le norme ("Regolamento") in base alle quali la RA Italiana opera sono definite dalla Naming Authority Italiana in base alle norme ISO 6523.
La Naming Authority Italiana [3] è l'organismo che stabilisce le procedure operative ed il regolamento in base al quale viene effettuata la registrazione dei nomi a dominio sotto il ccTLD ".it".
Oltre che per la gestione del TLD ".it" (nomi a dominio secondo lo standard Internet **RFC0822**), la RA è anche responsabile delle attività relative all'assegnazione di nomi definiti da altri standard (ITU X.400, ITU X.500).
Le attività della RA sono svolte dall'Istituto di Informatica e Telematica del Consiglio Nazionale delle Ricerche (IIT-CNR).
La RA fornisce anche servizi aggiuntivi, quali l'ospitalità sui sistemi della RA di nameserver secondari per quei domini per i quali ne sia fatta richiesta. Vengono inoltre mantenute varie liste di discussione su argomenti di interesse per la comunità Internet italiana, e vari database relativi a Internet, compreso ovviamente anche il database dei domini attivi nel Top Level Domain ".it".
Il ruolo di Registration Authority deriva al CNR dalla posizione che esso ricopre nella comunità scientifica nazionale ed internazionale quale Ente pubblico di ricerca. È stato affidato ai tecnici dello IIT con l'accordo dello IANA (Internet Assigned Number Authority) sulla base di riconosciute competenze acquisite dal suddetto personale che, a partire dalla metà degli anni '80, ha diffuso il protocollo IP nell'ambiente della ricerca italiana [4].
L'assegnazione dei nomi a dominio nel country code "IT" è anche affidata al GARR.
Il GARR [5], il cui acronimo originale significa "Gruppo per l'Armonizzazione delle Reti della Ricerca", è composto da tutte le Entità che rappresentano la Comunità Accademica e della Ricerca Scientifica in Italia, e si occupa della registrazione dei nomi a dominio esclusivamente per coloro che appartengono alla Comunità GARR o che hanno diritto di accesso alla rete GARR.
Il GARR infatti non è il gestore del ccTLD ".it", che è invece affidato alla Registration Authority Italiana.

WHOIS

Cosa è WHOIS

WHOIS è un programma che ricerca informazioni su utenti o domini della rete, interrogando un determinato database. Questo database contiene normalmente informazioni sui domini di secondo livello e sui relativi amministratori, per i quali fornisce dati che vanno dagli indirizzi e-mail agli indirizzi postali e ai numeri telefonici.
Può anche fornire informazioni circa le reti, le organizzazioni di networking, domini e siti.
Il principale database WHOIS è mantenuto dall' Internet Registration Service (InterNIC), ma parecchie organizzazioni gestiscono un loro database whois.
In realtà, i nomi dei contatti tecnici ed amministrativi per i domini registrati sono immessi automaticamente nel database quando le applicazioni di dominio o di indirizzo IP sono elaborate dalla *Internet coordination authority*. Ogni registrazione (entry) del database ha un handle (un identificatore univoco), un nome, un tipo di record, e vari altri campi dipendenti dal tipo di record.
Antecedentemente al 1° Aprile 1993 il Network Information Center (NIC) della Defense Data Network (DDN) era la Internet coordination authority e, perciò, mantenne il database (conosciuto come NIC database). Il NIC database è oggi ristretto alle informazioni circa il dominio **.mil**
Molti documenti si riferiscono ancora a questi nomi.

Ci sono delle limitazioni relative all'attuale implementazione di WHOIS che gli impediscono di diventare un servizio efficiente per un grande volume di informazioni e per richieste numerose: i vari server WHOIS non hanno conoscenza l'uno dell'altro, un database è mantenuto in ogni sito di server, infine, sono state implementate localmente e non propagate nuove funzionalità a vari siti. Per migliorare il servizio attuale sarà introdotto WHOIS++, un'evoluzione dell'attuale protocollo.

Chi può usare WHOIS

Ogni utente Internet può utilizzare WHOIS, o via TELNET (collegandosi ad un host che offre questo servizio) oppure interrogando il database whois Internic via web o anche via e-mail, inserendo la query nel corpo del messaggio, per esempio: "whois Giorgio Rossi".
Si possono ricercare informazioni che riguardano domini, host, reti, nomi ed indirizzi di posta.
Se si utilizza la connessione via TELNET si ricorda che il server whois è in ascolto sulla porta 43, ad esempio telnet *nomehost* 43 (anche se questa procedura è complessa da usare e sconsigliabile)
In generale si può dire che i server WHOIS andrebbero usati solo per interrogazioni specifiche, isolate e non per estrarre un numero elevato di informazioni.
Ricordiamo infine che, interrogare il database per estrarre dati utilizzabili per scopi commerciali è severamente proibito.

Alcuni database WHOIS

Inserendo un nome di dominio, come soggetto sul quale ricercare informazioni Whois restituirà il nome e l'indirizzo dei proprietari del dominio.
Dei diversi database WHOIS esistenti si ricordano:

- Il NSI Registrar database contiene domini e contatti non militari e non "US Government" (http://www.netsol.com);
- American Registry for Internet Numbers contiene informazioni circa Network, Autonomous System Numbers (ASN), ed i relativi Points of Contact (POC) (http://www.arin.net/whois/arinwhois.html);
- European IP Address allocations contiene i domini relativi all'Europa (http://www.ripe.net/db/whois.html);
- US Government contiene solo domini e contatti "US Government" (http://www.nic.gov/whois.html).

É possibile consultare i database della Registration Authority Italiana e di RIPE per reperire informazioni sulla registrazione di domini, persone, maintainer all'indirizzo: http://www.nic.it/RA/database/database.html.

I DNS nell'analisi del traffico web

Quando un server web crea un log file, tipicamente registra solo gli indirizzi IP dei visitatori. La maggior parte dei web server hanno la capacità di cercare il nome di dominio dei visitatori, ma poiché questa procedura diminuisce la velocità del server stesso, solitamente questa funzione viene disattivata. Anche nel caso in cui viene attivata, sempre per poter mantenere alte le prestazioni del server vengono impostati dei tempi di time-out decisamente limitati. Questo significa che il web server tenta di risolvere gli indirizzi, ma aspetta la risposta solo per poco tempo dopo il quale rinuncia e si accontenta dell'indirizzo numerico. Questo è il motivo della grossa percentuale di indirizzi non risolti solitamente presente nei file di log.
Per ovviare a questo fatto molti analizzatori di log file permettono di eseguire il DNS lookup prima del processo di analisi. Eseguire le statistiche in questo modo permette di visualizzare i nomi di dominio dei visitatori al posto dei loro indirizzi numerici che sarebbero privi di senso. Ovviamente attivare la risoluzione DNS è sempre molto dispendioso in termini di tempo. Per velocizzare questa procedura viene usata spesso una cache dove memorizzare tutti gli indirizzi di volta in volta risolti. Durante le ricerche successive prima di contattare il server DNS, l'indirizzo verrà cercato nella cache.
È importante osservare che non tutti gli indirizzi IP possono essere risolti con un nome di dominio. Questo accade nei seguenti casi:

1. Il server DNS non può essere contattato a causa di errori del server stesso o perché non raggiungibile;
2. All'indirizzo IP in questione non è stato associato nessun nome a dominio;
3. All'indirizzo IP è associato un nome di dominio attraverso il Resource Record di tipo A, ma potrebbe non essere definito il RR di tipo PTR corrispondente.

Bibliografia

[1] Mockapetris PV. RFC1034, D*omain names – concepts and facilities*, novembre1987.
Disponibile all'indirizzo URL:
http://isc.faqs.org/rfcs/rfc1034.html
Data ultimo accesso: 11 aprile 2003.

[2] Registration Authority Italiana. Home Page.
Disponibile all'indirizzo URL:
http://www.nic.it/RA/index.html.
Data ultimo accesso: 11 aprile 2003.

[3] Naming Authority Italiana. Home Page.
Disponibile all'indirizzo URL:
http://www.nic.it/NA/.
Data ultimo accesso: 11 aprile 2003.

[4] Registration Authority Italiana.
Informazioni generali.
Disponibile all'indirizzo URL:
http://www.nic.it/RA/info/info.html.
Data ultimo accesso: 11 aprile 2003.

[5] GARR. Registrazione di nomi a dominio. Disponibile all'indirizzo URL:
http://www.garr.it/garr-b-domini.shtml.
Data ultimo accesso: 11 aprile 2003.