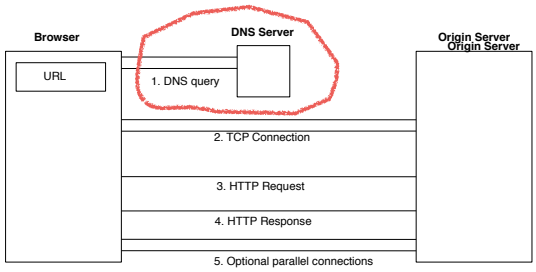
Τα ονόματα στο ΠΠΠ



Domain Name System (DNS)

- a set of servers that map written names to IP addresses
  - Example: www.cs.washington.edu → 128.208.3.88
- many systems maintain a local cache called a hosts file
  - Windows: C:\Windows\system32\drivers\etc\hosts
  - Mac: /private/etc/hosts
  - Linux: /etc/hosts

DNS: Outline

- Computer science concepts underlying DNS
  - Indirection: names in place of addresses
  - Hierarchy: in names, addresses, and servers
  - Caching: of mappings from names to/from addresses
- Inner-workings of DNS
  - DNS resolvers and servers
  - Iterative and recursive queries
  - TTL-based caching
- Web and DNS
  - Influence of DNS queries on Web performance
  - Server selection and load balancing

Host Names vs. IP addresses

- Host names
  - Mnemonic name appreciated by humans
  - Variable length, alpha-numeric characters
  - Provide little (if any) information about location
  - Examples: www.cnn.com and ftp.eurocom.fr
- IP addresses
  - Numerical address appreciated by routers
  - Fixed length, binary number
  - Hierarchical, related to host location
  - Examples: 64.236.16.20 and 193.30.227.161

Separating Naming and Addressing

- Names are easier to remember
  - www.cnn.com vs. 64.236.16.20
- Addresses can change underneath
  - Move www.cnn.com to 64.236.16.20
  - E.g., renumbering when changing providers
- Name could map to multiple IP addresses
  - www.cnn.com to multiple replicas of the Web site
- Map to different addresses in different places
  - Address of a nearby copy of the Web site
  - E.g., to reduce latency, or return different content
- Multiple names for the same address
  - E.g., aliases like ee.mit.edu and cs.mit.edu

Strawman Solution: Local File

- Original name to address mapping
  - Flat namespace
  - /etc/hosts
  - SRI kept main copy
  - Downloaded regularly
- Count of hosts was increasing: moving from a machine per domain to machine per user
  - Many more downloads
  - Many more updates

Strawman Solution #2: Central Server

- Central server
  - One place where all mappings are stored
  - All queries go to the central server
- Many practical problems
  - Single point of failure
  - High traffic volume
  - Distant centralized database
  - Single point of update
  - Does not scale

Need a distributed, hierarchical collection of servers

Domain Name System (DNS)

- Properties of DNS
  - Hierarchical name space divided into **zones**
  - Distributed over a collection of **DNS servers**
- Hierarchy of DNS servers
  - Root servers
  - Top-level domain (TLD) servers
  - Authoritative DNS servers
- Performing the translations
  - Local DNS servers
  - Resolver software

DNS - The Internet Domain Name System

- A distributed naming database
- Name structure reflects administrative structure of the Internet
- Rapidly resolves domain names to IP addresses
  - exploits caching heavily
  - typical query time ~100 milliseconds
- Scales to millions of computers
  - partitioned database
  - caching
- Resilient to failure of a server
  - replication

Αρχιτεκτονική DNS

- **Ζώνες** (zones): μη επικαλυπτόμενα τμήματα ενός χώρου ονομάτων, το καθένα εκ των οποίων υποστηρίζεται από διαφορετικό εξυπηρετητή
  - υποδένδρα της ιεραρχίας του DNS, τα οποία ανήκουν σε διαφορετική διοικητική αρχή.
  - Κάθε ζώνη έχει συνήθως έναν πρωταρχικό εξυπηρετητή και περισσότερους δευτερεύοντες εξυπηρετητές, οι οποίοι μπορούν να υποκαταστήσουν τον πρωταρχικό σε περίπτωση βλάβης.
  - Μεγάλοι οργανισμοί μπορούν να οργανώνουν τα πεδία τους σε περισσότερες της μιας ζώνες.
- Η αποδοτική απεικόνιση διευθύνσεων IP σε ονόματα κόμβων (hostnames) προϋποθέτει μια διαφορετική ιεραρχία, βασιζόμενη σε διευθύνσεις IP.
  - Η ανάθεση των διευθύνσεων IP υποστηρίζεται από μητρώα (registries): APNIC (Ασία), ARIN (Β. Αμερική), RIPE NCCC (Ευρώπη), LACNIC, AFRINIC
- Η ανάθεση διευθύνσεων IP στα τρία μητρώα γίνεται από το **Internet Assigned Numbers Authority (IANA)**, που είναι τμήμα του μη κερδοσκοπικού οργανισμού **ICANN (Internet Corporation for Assigned Names and Numbers)**.
  - Με την δέσμευση ενός συνόλου διευθύνσεων IP από έναν οργανισμό, ο οργανισμός αυτός γίνεται υπεύθυνος για ένα τμήμα του ονοματοχώρου in-addr.arpa. Πρόκειται για μια ιεραρχία που βασίζεται στις οκτάδες των 32-μπιτων διευθύνσεων IP.

THE WALL STREET JOURNAL.

Subscribe Now | Sign In

SPECIAL OFFER: JOIN NOW

Home World U.S. Politics Economy Business Tech Markets Opinion Arts Life Real Estate

Search

Samsung to Recall Galaxy Note 7 Over Reports of Fires

Andreessen Firm's Returns Trail Venture-Capital Elite

Rocket Explosion Leaves Facebook's Internet Initiative Grounded

China Opens Antitrust Probe of Didi-Uber Deal

TECH

Obama Administration to Privatize Internet Governance on Oct. 1

Transfer of domain-name authority from U.S. likely to spark debate in Congress

By JOHN D. MCKINNON

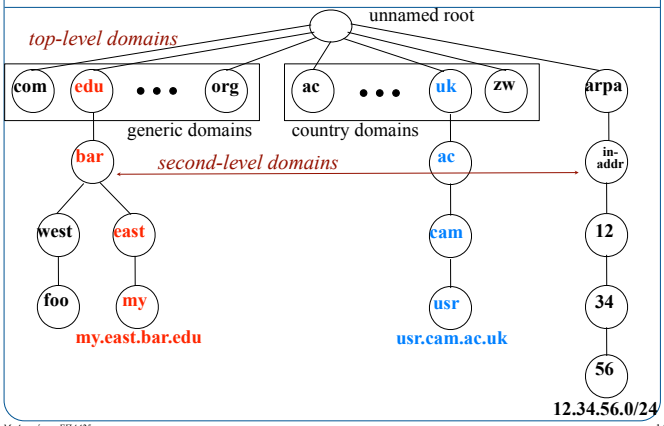
Updated Aug. 16, 2016 5:25 p.m. ET

WASHINGTON—The Obama administration said Tuesday it will formally shift authority for much of the internet's governance to a nonprofit multi-stakeholder entity on Oct. 1, a move likely to spark a backlash from parts of Congress.

TLD and Authoritative DNS Servers

- Top-level domain (TLD) servers
  - Generic domains (e.g., com, org, edu)
  - Country domains (e.g., uk, fr, ca, jp)
  - Typically managed professionally
    - Network Solutions maintains servers for “com”
    - Educause maintains servers for “edu”
- Authoritative DNS servers
  - Provide public records for hosts at an organization
  - For the organization’s servers (e.g., Web and mail)
  - Can be maintained locally or by a service provider

Distributed Hierarchical Database

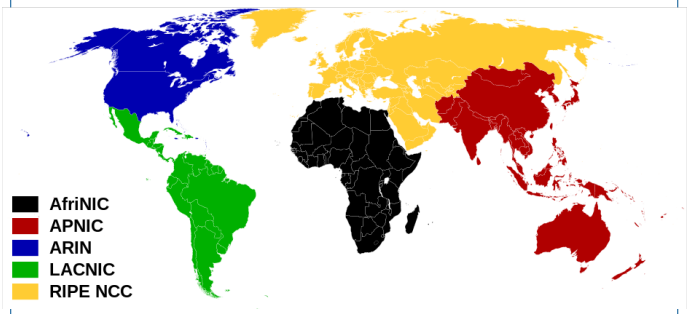


The role of root servers

Notice that:

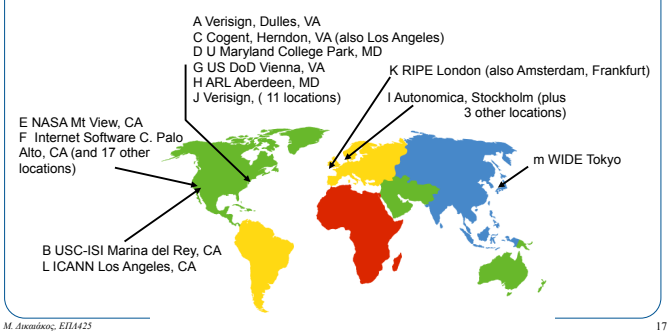
- root servers only know who you need to ask next.
  - .com -> list of servers
  - .net -> list of servers
  - .ch -> list of servers
  - .ug -> list of servers
  - .br -> list of servers
- Caching of previous answers means there is less need to query the root servers after the first question.

http://www.root-servers.org/presentations/wsis.pdf



DNS Root Servers

- 13 root servers (see http://www.root-servers.org/)
- Labeled A through M



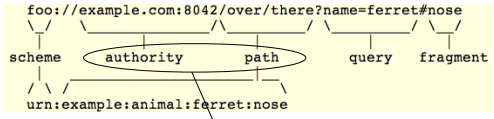
Universal Resource Identifiers

Universal Resource Identifier

- Προδιαγραφή ονοματοδοσίας ιστιακών πόρων
- Υποστηρίζει ένα γενικό τρόπο για ανάθεση ονομάτων σε πόρους του Ιστού
- Υιοθετεί ένα ομοσπονδιακό, ιεραρχικό και επεκτάσιμο συντακτικό για την δημιουργία ονομάτων
  - Με βάση το συντακτικό του, κάθε URI ξεκινάει με το όνομα ενός σχήματος (schema) στο οποίο ανήκει το συγκεκριμένο URI
  - Η δομή του υπολοίπου URI ακολουθεί σύνταξη που είναι συγκεκριμένη για το συγκεκριμένο σχήμα

Syntax Components

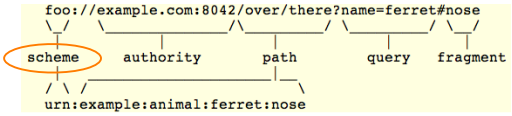
- URI syntax consists of a hierarchical sequence of components referred to as:
  - Scheme, Authority, Path, Query, Fragment




```
URI         = scheme ":" hier-part [ "?" query ] [ "#" fragment ]

hier-part   = "//" authority path-abempty
            / path-absolute
            / path-rootless
            / path-empty
```


URI Scheme

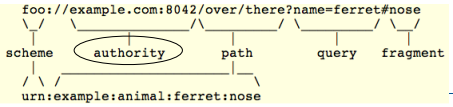


- The URI scheme refers to a specification for assigning identifiers *within* a scheme.
- The Internet Assigned Numbers Authority (IANA) maintains a registry of mapping between URI scheme names and their specifications.
- The cost of introduction of new URI schemes is high (why?)

URI Authority part

- Many URI schemes include a hierarchical element for a naming authority so that governance of the name space defined by the remainder of the URI is delegated to that authority. The owner of a resource assigns:
  - The authoritative meaning of the resource.
  - The URIs that refer to the resource.
- Within a URI we can distinguish the authority based on: a registered name or server address, along with optional port and user information:
  - The authority component is preceded by a double slash ("//") and is terminated by the next slash ("/"), question mark ("?"), or number sign ("#") character, or by the end of the URI.
  - If a URI contains an authority component, then the path component must either be empty or begin with a slash ("/") character.

```
authority   = [ userinfo "@" ] host [ ":" port ]
```

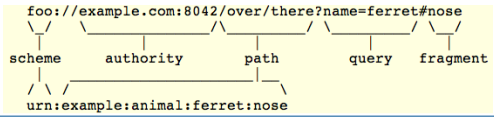


URIs, URLs and URNs

- A URI can be further classified as a locator, a name, or both.
  - "Uniform Resource Locator" (URL): refers to the subset of URIs that, in addition to identifying a resource, provide a means of locating the resource by describing its primary access mechanism (e.g., its network "location").
  - "Uniform Resource Name" (URN): has been used historically to refer to both URIs under the "urn" scheme, which are required to remain globally unique and persistent even when the resource ceases to exist or becomes unavailable, and to any other URI with the properties of a name.
  - An individual scheme does not have to be classified as being just one of "name" or "locator".

Path component

- Contains data, usually organized in hierarchical form that serves to identify a resource within the scope of the URI's scheme and naming authority.
- The path is terminated by the first question mark ("?") or number sign ("#") character, or by the end of the URI.
  - If a URI contains an authority component, then the path component must either be empty or begin with a slash ("/") character.
  - If a URI does not contain an authority component, then the path cannot begin with two slash characters ("//").
  - A URI reference may be a relative-path reference, in which case the first path segment cannot contain a colon (":") character.



Examples

ftp://ftp.is.co.za/rfc/rfc1808.txt

http://www.ietf.org/rfc/rfc2396.txt

ldap://[2001:db8::7]/c=GB?objectClass?one

mailto:John.Doe@example.com

news:comp.infosystems.www.servers.unix

tel:+1-816-555-1212

telnet://192.0.2.16:80/

urn:oasis:names:specification:docbook:dtd:xml:4.1.2

- **Note: separation of identification from interaction**

Query component

- The query component contains non-hierarchical data that serves to identify a resource within the scope of the URI's scheme and naming authority.
- The query component is indicated by the first question mark ("?") character and terminated by a number sign ("#") character or by the end of the URI.
- Query components are often used to carry identifying information in the form of "key=value" pairs.

Fragment Identifiers

- The fragment identifier component of a URI allows indirect identification of a *secondary resource*, by reference to a primary resource and additional identifying information that is selective with respect to that resource. The identified secondary resource may be:
  - some portion or subset of the primary resource
  - some view on representations of the primary resource,
  - some other resource that is merely named with respect to the primary resource.
- Some URI schemes do not specify the use of fragment identifiers (e.g. MAILTO URIs).
- For URI schemes that do specify the use of fragment identifiers, the syntax and semantics of those identifiers is defined by the set of representations that might result from a retrieval action on the primary resource. The presence of a fragment identifier component in a URI does not imply that a retrieval action will take place.
- Interpretation of the fragment identifier during a retrieval action is performed solely by the user agent; the fragment identifier is not passed to other systems during the process of retrieval.

Relative URIs

- The URI spec. allows scheme-independent references to be made relative to a hierarchy.
  - It is often the case that a group or "tree" of documents has been constructed to serve a common purpose, and most URI references in these documents point to resources within the tree.
  - Relative referencing of URIs allows document trees to be partially independent of their location and access scheme.
  - Furthermore, such document trees can be moved as a whole, without changing any of the relative references.
  - A relative reference refers to a resource by describing the difference within a hierarchical name space between the reference context and the target URI.
  - A reference resolution algorithm, defines how such a reference is transformed to the target URI.

Interactions through a URI

- Resolution (ανάλυση/επίλυση)
  - The process of determining an access mechanism and the appropriate parameters necessary to dereference a URI; such resolution may require several iterations.
- Dereference (πρόσβαση)
  - To dereference a URI is to use an access mechanism to perform an action on the URI's resource.
- Retrieval (ανάκτηση)
  - A URI dereference that causes an agent to retrieve a representation of the associated resource.

Resolution and Retrieval

- During URI resolution, an agent applies in succession a finite set of relevant specifications, beginning with the specification of the context in which the URI is found (e.g., a format or protocol specification, or an application). Any one of these specifications may define more than one access mechanism.
- One of the most important actions on the Web is to retrieve a representation of a resource (for example, by using HTTP GET).
  - As stated above, the authority responsible for a URI determines what the URI identifies and which representations are used for interaction with the resource.
  - The representations communicate the meaning of the resource.

Uniform Resource Locator (URL)

- RFC1738
- an identifier (αναγνωριστικό / όνομα) for the location of a document on a web site
- a basic URL:

http:// www.aw-bc.com /info/regesstepp/index.html

protocol host path

- upon entering this URL into the browser, it would:
  - **ask the DNS server** for the IP address of www.aw-bc.com
  - connect to that IP address at port 80 (**open socket**)
  - ask the server to **GET** /info/regesstepp/index.html
  - **display** the resulting page on the screen

Πιο σύνθετα URL

- **anchor (άγκυρα)**: jumps to a given section of a web page
  http://www.textpad.com/download/index.html**#downloads**
  - fetches index.html then jumps down to part of the page labeled downloads
- **port (θύρα)**: for web servers on ports other than default 80
  http://www.cs.washington.edu**:8080**/secret/money.txt
- **query string (αλφαριθμητικό επερώτησης)**: a set of parameters passed to a web program
  http://www.google.com/search**?q=miserable+failure&start=10**
  - parameter q is set to "miserable+failure"
  - parameter start is set to 10

Στοιχεία Συντακτικού URL

- Σχήμα (scheme, followed by a colon)
  - http:, ftp:, mailto:, telnet:
- // (Double slash (only for http, ftp, wais, gopher)
- Όνομα πεδίου (Internet domain name) ή διεύθυνση IP
- Αριθμός θύρας προαιρετικά (port number) - eg. www.cs.ucy.ac.cy:8080
  - Default ports:
    - HTTP is 80
    - FTP is 21
    - SMTP is 25
    - IMAP is 143
- Μονοπάτι (path) - e.g. /users/mdd