                   Domain Names and Company Name Retrieval

Status of this Memo

   This memo defines an Experimental Protocol for the Internet
   community.  It does not specify an Internet standard of any kind.
   Discussion and suggestions for improvement are requested.
   Distribution of this memo is unlimited.

Copyright Notice

Abstract

   Location of web information for particular companies based on their
   names has become an increasingly difficult problem as the Internet
   and the web grow.   The use of a naming convention and the domain
   name system (DNS) for that purpose has caused complications for the
   latter while not solving the problem.  While there have been several
   proposals to use contemporary, high-capability, directory service and
   search protocols to reduce the dependencies on DNS conventions, none
   of them have been significantly deployed.

   This document proposes a company name to URL mapping service based on
   the oldest and least complex of Internet directory protocols, whois,
   in order to explore whether an extremely simple and widely-deployed
   protocol can succeed where more complex and powerful options have
   failed or been excessively delayed.

1. Introduction and Context

   In recent months, there have been many discussions in various
   segments of the Internet community about "the top level domain
   problem".  Perhaps characteristically, that term is used by different
   groups to identify different, and perhaps nearly orthogonal, issues.
   Those issues include:

1.1.  A "domain administration policy" issue.

1.2.  A "name ownership" issue, of which the trademark issue may
      constitute a special case.

1.3.  An information location issue, specifically the problem of
      locating the appropriate domain, or information tied to a
      domain, for an entity given the name by which that entity is
      usually known.

Of these, controversies about the first two may be inevitable
consequences of the growth of the Internet.  There have been
intermittent difficulties with top level domain adminstration and
various attempts to use the domain registry function as a mechanism
for control of service providers or services from time to time since
a large number of such domains started being allocated.  Those
problems led to the publication of the policy guidelines of
[RFC1591].

The third appears to be largely a consequence of the explosive growth
of the World Wide Web and, in particular, the exposure of URL formats
[URL] to the end user because no other mechanisms have been
available.  The absence of an appropriate and adequately-deployed
directory service has led to the assumption that it should be
possible to locate the web pages for a company by use of a naming
convention involving that company's name or product name, i.e., for
the XYZ Company, a web page located at

     http://www.xyz.com/
or
     http://www.xyz-company.com/

has been assumed.

However, as the network grows and as increasing numbers of web sites
are rooted in domains other than ".COM", this convention becomes
difficult to sustain: there will be too many organizations or
companies with legitimate claims --perhaps in different lines of
business or jurisdictions-- to the same short descriptive names.  For
that reason, there has been a general sense in the community for
several years that the solution to this information location problem
lies, not in changes to the domain name system, but in some type of
directory service.

But such directory services have not come into being.  There has been
ongoing controversy about choices of protocols and accessing
mechanisms.  IETF has published specifications for several different
directory and search protocols, including [WHOIS++], [RWHOIS],

[LDAP], [X500], [GOPHER].  One hypothesis about why this has not
happened is that these mechanisms have been hard to select and deploy
because they are much more complex than is necessary.  This document
proposes an extremely simple alternative.

2. Using WHOIS

The WHOIS protocol is the oldest directory access protocol in use on
the Internet, dating in published form to March 1982 and first
implemented somewhat earlier.  The procotol itself is simple and
minimalist: the client opens a telnet connection to the WHOIS port
(43) and transmits a line over it.  The server looks up the line in a
fashion that it defines, returns one or more lines of information to
the client, and closes the connection.

We suggest that modifications or add-ins be created to Web browsers
that would access a new, commercially-provided Whois server, sending
a putative company name and receiving back one or more lines, each
containing a URL followed by one or more blanks and then a matching
company name (that order was chosen to minimize parsing problems:
since URLs cannot contain blanks, the first blank character marks the
end of the URL and the next non-blank marks the beginning of the
company name).  As is usual with Whois, the criteria used by the
server to match the incoming string is at the server's discretion.
The difference between this and the protocol as documented in [WHOIS]
is that exactly one company name is returned per line (see section 3
for details of syntax).

The client would then be expected to:

(i) If a single line (company name and URL) is returned, either
    ask for confirmation or simply fetch the associated URL as if it
    had been typed by the user.

(ii) If multiple lines (names) are returned, present the user with
     a choice, presumably showing company names rather than (or
     supplemented by) URLs, then fetch using the URL selected.

Obviously, while the most convenient use of the services contemplated
in this document would occur through a client that was part of, or
intimately connected with, a Web browser, a user without that type of
facility could utilize a traditional WHOIS client and paste or
otherwise transfer the relevant information into the target location
of a browser.

3. Formats, versions, and international character sets

   Preliminary work with the approach suggested above suggests that some
   specific conventions about syntax and variations would be useful.

3.1 Line sent from client to server.

   These lines may take either of two forms:

   (i) A simple 7-bit ASCII string, containing a "company name"

   (ii) A string in the format (using the ABNF notation of RFC 2234
        [ABNF]):

        Variation "/" 1*Octet

            Variation :== "0" | ( Non-zero-digit 1*Digit)
            Non-zero-digit :== 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9
            Digit :== 0 | Non-zero-digit

        Where Octet is any eight-bit sequence, representing a prefixed
        variation number.

   The first form will be construed as equivalent to the second form
   with the leading string "0/".  Variation numbers are specified in
   section 3.3.

   In all cases, the interpretation of what "company name" might mean
   and, in particular, what variations of form or spelling,
   abbreviations, and so on, might be accepted is strictly up to the
   interpretation of the server.  If rules driving the server lead to
   the conclusion that a string matches some company in its data, the
   correctness or incorrectness of that decision is not covered by this
   specification.

   For variation 0 and, by default, for all others, any alphabetic text
   in lines is to be construed in a case-insensitive fashion.

3.2 Lines sent from server to client.

   The server is expected to return one or more lines to the client,
   depending on its interpretation of the input string.  In general,
   each line will consist, as described above, of a URL, a space, and a
   "company name".  This document deliberately does not specify the
   content or semantics of the "company name" string.  It might be a
   name, or a name and descriptive information such as location and type
   of business, or other information at the option of the server.  The

expectation, as mentioned above, is that the information will be
displayed by the client to aid users in selecting the appropriate
URL.

These lines, consistent with normal Internet practice, will be
terminated by a CR LF sequence (rather than one or the other of those
control characters).

When and if different variation numbers are introduced, their
specifications may include variations on what the server is expected
to return.

In lieu of "URL and company name" responses, the Server may also
return "error messages".  These take the form of lines containing:

        "///" SP String

 where the String is 7-bit ASCII with no control characters other
 than SP, unless the variation associated with the variation number
 specifies otherwise.  For this experiment, all "error messages" but
 the following two are discouraged:

        /// Not found
                Indicating that the "company name" does not match
                anything
        /// Variation not supported
                Indicating that the variation number supplied by the
                client is not recognized by the server.

3.3.  Registered variations

   The following two variations are established as part of this
   specification:

   0/        Query and response are in 7-bit ASCII, no controls other
             than SP, "Company name" separated from URL by one or more
             SP characters.

   1/        Query and response are in UTF-8, no controls other than
             SP, "Company name" separated from URL by one or more SP
             characters, no specification of language on either input or
             output.

   The IANA will maintain a registry of additional variations which it
   is hoped will be very short.  Requests for additional variations
   should be sent via email to: iana@iana.org.

4. Alternatives not chosen

   Few comments on the initial drafts of this document addressed the
   basic model or protocol design for the service discussed.  Instead,
   they focused on inquiring about the decisions we didn't make and
   about beliefs about the protocol specification that were not intended
   by the authors.  The latter have been, we hope, corrected.  Questions
   of the following three types predominated in the first category.

4.1.  Why didn't you use <insert-favorite-directory-protocol-here>?

   Many notes raised the question of how much more could be done with a
   higher-powered directory protocol rather than the extremely simple
   WHOIS.  Questions were raised about LDAP, X.500 DAP, CCSO, RWHOIS,
   and WHOIS++.  We had several reasons for avoiding them.  The most
   important has been a strong commitment to see how much can be done
   with an extremely simplistic approach, and WHOIS represented the most
   simplistic approach we could find.  If it turns out to be too simple
   in practice, things can always evolve to one or more of the more
   advanced protocols.   But, if we started with one of them, we would
   never get that information.  Other issues included:

   * None of the existing directory proposals has really emerged as
     the "right" solution with a large installed base.  The deployed
     base of WHOIS and WHOIS clients is huge, and using it avoids either
     having to make a premature choice of "winner" or to become
     embroiled in the debate.

   * For the casual user, the mechanisms needed to activate the
     extensive attribute-based directory searches of the stronger
     protocols are just too complicated and may actually act as a
     deterrent to effective use.

   * Substantially since the dawn of the ARPANET, the Internet
     experience has been that setting up a directory service is easy,
     but that maintaining one and keeping the records up-to-date is
     extremely difficult.  The economics of operating an effective
     directory service and keeping everything up to date may will
     require a revenue-producing product.  Use of a very simple protocol
     for the basic service creates a situation in which basic service
     can rationally be given away while more advanced service are
     operated on a charge or subscription basis.

4.2 And why not use a Web search engine?

   Web search engines are immensely effective and powerful, but address
   a different problem than this protocol.  The protocol model here does
   involve a directory lookup, using a presumed company name as a key.

The quality of the result will depend on the quality of the
underlying directory and the editorial and research work that goes
into its construction (neither of which are matters for the protocol
itself -- we trust that marketplace pressures will separate good
servers from poor ones).  Web search engines are often more effective
at locating information about companies than the specific company-
designated web pages.

4.3 Why not return a more highly structured information format
     rather than a simple pair of URL and "company name"?

Again, the goal was to keep things extremely simple and, in
particular, permit minimal interpretation between the user's input
and the query and between the response and a display or action.  Some
of the inquiries on this subject were due to misunderstandings about
the implications of the "company name" field; the semantics of that
field have been clarified above.  We also wanted to avoid the level
of standardization implied by a tagging scheme: highly-structured
fields might lead either to interoperability problems or excessive
restriction on what might be returned.

5. Thoughts on Directory Providers

There is no technical reason why there should be only one provider of
company name to URL mapping services using this protocol, nor is
there any reason for registries of such providers.  Presumably,
servers that provide the best-quality mappings will eventually
prevail in the marketplace.  However, as with most traditional uses
of WHOIS, it is desirable for implementations of clients (or Web
browsers supporting this protocol) to allow for user choice of
servers through configuration options or the equivalent.

6. Demo Application

To illustrate the proposed functionality of this document, a
prototype of both the server and client have been made able for
demonstration purposes.

6.1 Server

The TLD-WHOIS demonstration server is available at
"companies.mci.net". The server contains a database of approximately
209,000 company entries provided by Dun and Bradstreet.

The server will generally respond back to a query within 15 seconds.
If the server has the response cached from a previous query, the
return time will be significantly shorter.

   If 10 or more entries are found in the database for the query, only
   the top 10 will be returned in the response.

   For the purposes of this demonstration, there is no provision for
   submitting additions or changes to the database. The authors and the
   sponsoring companies are not responsible for the accuracy of the data
   provided by this prototype. Our apologies if your company is not
   listed.

6.2  Client

6.2.1 Download Location:

   A demonstration client for the Windows 95/Nt platforms is available
   for public download through anonymous ftp at:
   ftp.mci.net/pub/ietf/company/demo.exe, or via the web:
   ftp://ftp.mci.net/pub/ietf/company/demo.exe
   File size is approximately 1.9 MB.

6.2.2 Setup Instructions:

   a) Download the client installation software from the site mentioned
      above to a local 32 bit Windows computer. The client installation
      software has been compressed using the self-extracting archive
      application from InstallShield The default name for the download
      is "demo.exe".

   b) Double click on the file through File Explorer or run the program
      through the START menu.

   c) Select "Setup" to allow InstallShield to uncompress the files
      needed to install the demonstration client to a temporary
      directory. InstallShield will then automatically launch the main
      application Setup program.

   d) The main setup program will install the demo application files and
      make the necessary additions to the Windows Registry. No user
      action is required.

   e) Upon completion of installation you will be prompted to run the
      application or to exit setup.

6.2.3 Paranoia:

   What did you just do to my computer?

   Files Copied:

   companyname.exe    Main program executable
   whois.ocx          WhoIs module from Mabry Software
   led.ocx            LED module from Mabry Software
   msvbvm50.dll       Microsoft Visual Basic 5.0 runtime file
   stdole2.tlb        Microsoft Visual Basic 5.0 runtime file
   oleaut32.dll       Microsoft Visual Basic 5.0 runtime file
   olepro32.dll       Microsoft Visual Basic 5.0 runtime file
   comcat.dll         Microsoft Visual Basic 5.0 runtime file
   asyncfilt.dll      Microsoft Visual Basic 5.0 runtime file
   crtl3d32.dll       Installshield control used for installation only

   Registry Changes:

   Created key under HKEY_CLASSES_ROOT called Who

   This entry is used to enable the Microsoft Internet Explorer's
   pluggable protocol handler. The key contains several sub-entries that
   list the path and command to the companyname executable. The
   pluggable protocol hander provides the necessary hooks to launch the
   companyname application whenever the WHO:// URL is submitted in the
   address line of Internet Explorer.

6.2.4 Using the Program

6.2.4.1 Standalone Operation:

   From the Start Menu, select the Programs \ Companyname \ companyname.
   Alternatively, it can be launched from Start:
     Run c:\windows\companyname.exe

   Enter the name of the company that you are attempting to locate and
   press OK.

   A status box will be displayed while the client is communicating with
   the server until a response is returned. The possible returns are:

     a) Message box saying that,  "Your request was not found."
        This means that the company information that was submitted was
        not found in the database.

      b) A list box containing 2 - 10 company names sorted high to
         low by score. Highlight one of the names and press the launch
         button. The program will launch the default web browser for
         your computer and navigate to the site.

      c) The default web browser launches and navigates to a site.
         This means that only one match was found in the database and
         that match is opened directly without user intervention.

6.2.4.2 Within Internet Explorer

   From the Address Line within the web browser, enter "WHO://" followed
   by the name of the company that you wish to search for and press the
   enter key.

      Note:  Since the company name is entered within the URL space
             of the browser, it can not contain spaces.

   If you wish to send a search string that contains spaces, enter
   "WHO://" with no company information.  The application will display
   the dialogue window as described in standalone mode for you to enter
   the search criteria.

   A status box will be displayed while the client is communicating with
   the server until a response is returned. The possible returns are:

      a) Message box saying that,  "Your request was not found."
         This means that the company information that was submitted was
         not found in the database.

      b) A list box containing 2 - 10 company names sorted high to
         low by score. Highlight one of the names and press the launch
         button. The program will launch the default web browser for
         your computer and navigate to the site.

      c) The default web browser launches and navigates to a site.
         This means that only one match was found in the database and
         that match is opened directly without user intervention.

6.2.5 Client Customization

   The name of the Whois server is hardcoded within the application to
   "companies.mci.net". No initialization file or registry keys are
   needed for the default configuration.  Realizing  that some testers
   may have proxy servers on their corporate systems and that others may
   wish to test the client against a different Whois server, the client
   supports a mechanism for changing the default server.  To enable the
   server customization, follow these steps:

   a) Create a new directory in the root of the
      C: Drive called "companyname"

   b) Using Notepad or any text editor create a new file
      called "whois.ini"

   c) Add a new line to the file beginning with
      "SERVER= <server name>". Do not include the double quotes
      around the tag. <server name> would be the IP Address or DNS
      name of the new Whois or proxy server.

   d) End the line with a carriage return.

   e) Save the file as a plain text file back to
      "c:\companyname\whois.ini"

6.2.6 Client Limitations:

   The demonstration software and database are provided "as is". No
   warranties are stated or implied. Use at your own risk.

   The demonstration client is supported only on 32 bit Intel Windows
   platforms. It has been tested on  Windows 95, Windows NT 4.0 and
   Windows 98 beta RC0.

   Use of the WHO:// URL moniker from within the web browser is
   supported only under Microsoft Internet Explorer.

   TCP Port 43 must be cleared through firewalls for client to
   communicate with the server. Refer to the section on client
   customization if you need to utilize a proxy server to traverse a
   firewall.

   When using the Address Line entry method within Microsoft Internet
   Explorer, spaces are not permitted within the search string.

7. References

   [ABNF]  Crocker, D., and P. Overell, Eds., "Augmented BNF for Syntax
   Specifications: ABNF",  RFC 2234, November 1997.

   [RFC1591]  Postel, J., "Domain Name System Structure and Delegation",
   RFC 1591, March 1994.

   [GOPHER] Anklesaria, F., McCahill, M., Lindner, P., Johnson, D.,
   John, D., Torrey, D., and B. Alberti, "The Internet Gopher Protocol
   (a distributed document search and retrieval protocol)", RFC 1436,
   March 1993.

   [LDAP]  Yeong, W., Howes, T., and S. Kille, "Lightweight Directory
   Access Protocol", RFC 1777, March 1995.

   [RWHOIS]   Williamson, S., and M. Kosters, "Referral Whois Protocol
   (RWhois)", RFC 1714, December 1994.

   [URL]   Berners-Lee, T., Masinter, L., and M. McCahill, "Uniform
   Resource Locators (URL)", RFC 1738, December 1994.

   [WHOIS] Feinler, E., Harrenstien, K., and M. Stahl, "NICNAME/WHOIS",
   RFC 954, October 1985.

   [WHOIS++]  Deutsch, P., Schoultz, R., Faltstrom, P., and C. Weider,
   "Architecture of the WHOIS++ service", RFC 1835, August 1995.

   [X500]  Wright, R., Getchell, A., Howes, T., Sataluri, S., Yee, P.,
   and W. Yeong, "Recommendations for an X.500 Production Directory
   Service", RFC 1803, June 1995.

   [Z39.50]  Lynch, C., "Using the Z39.50 Information Retrieval Protocol
   in the Internet Environment", RFC 1729, December 1994.

8. Security Considerations

   This suggested use of the WHOIS protocol adds no significant security
   risks to those of traditional applications of the protocol which is
   one of the most widely-deployed applications on the Internet.  As
   usual, servers should expect to use the string sent to them as an
   information retrieval key, not as a function to be executed in some
   way.  A more significant risk would arise if the server supporting
   the translation function were somehow spoofed; in that case, an
   incorrect URL might be returned for a particular company. As with the
   possibility of finding an incorrect page using naming conventions,
   the best protection against the risks that could then occur is
   careful attention to certificates, signatures, and other
   authenticity-indicating information.

9.  IANA Considerations

   As provided in section 3.3, above, this experiment requests that IANA
   maintain a registry of query variation forms and that the registry be
   initialized with the two values specified in that section.

10. Acknowledgements

   This memo was inspired by a many discussions over the last few years
   about the status and uses of the domain name system, information
   location using conventions about domain names, exposure of URLs to
   end users, and convergence of directory and search protocols.  While
   the people involved are too numerous to attempt to list, the authors
   would like to acknowledge their contributions and comments.

   Martin Hamilton, Keith Moore, Tom Thornbury and Ed Trembicki-Guy made
   important suggestions that have contributed to the revision of this
   memo.

11. Authors' Addresses

   John C. Klensin
   MCI Internet Architecture
   800 Boylston St, 7th floor
   Boston, MA 02199
   USA

   Phone: +1 617 960 1011
   EMail: klensin@mci.net


   Ted Wolf, Jr.
   Electronic Commerce
   Dun & Bradstreet Information Services
   3 Sylvan Way
   Parsippany, NJ 07054
   USA

   Phone: +1 201 605 6308
   EMail: ted@usa.net


   Gary W. Oglesby
   MCI Internet Architecture
   842 N. Ahoy Dr.
   Gilbert, AZ 85234
   USA

   Phone: +1 415 538 1100
   EMail: gary@mci.net

12.  Full Copyright Statement