          A DNS RR for specifying the location of services (DNS SRV)

Status of this Memo

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

Abstract

   This document describes a DNS RR which specifies the location of the
   server(s) for a specific protocol and domain (like a more general
   form of MX).

Overview and rationale

   Currently, one must either know the exact address of a server to
   contact it, or broadcast a question. This has led to, for example,
   ftp.whatever.com aliases, the SMTP-specific MX RR, and using MAC-
   level broadcasts to locate servers.

   The SRV RR allows administrators to use several servers for a single
   domain, to move services from host to host with little fuss, and to
   designate some hosts as primary servers for a service and others as
   backups.

   Clients ask for a specific service/protocol for a specific domain
   (the word domain is used here in the strict RFC 1034 sense), and get
   back the names of any available servers.

Introductory example

   When a SRV-cognizant web-browser wants to retrieve

      http://www.asdf.com/

   it does a lookup of

      http.tcp.www.asdf.com

   and retrieves the document from one of the servers in the reply. The
   example zone file near the end of the memo contains answering RRs for
   this query.

The format of the SRV RR

   Here is the format of the SRV RR, whose DNS type code is 33:

        Service.Proto.Name TTL Class SRV Priority Weight Port Target

        (There is an example near the end of this document.)

   Service
        The symbolic name of the desired service, as defined in Assigned
        Numbers or locally.

        Some widely used services, notably POP, don't have a single
        universal name. If Assigned Numbers names the service
        indicated, that name is the only name which is legal for SRV
        lookups. Only locally defined services may be named locally.
        The Service is case insensitive.

   Proto
        TCP and UDP are at present the most useful values
        for this field, though any name defined by Assigned Numbers or
        locally may be used (as for Service). The Proto is case
        insensitive.

   Name
        The domain this RR refers to. The SRV RR is unique in that the
        name one searches for is not this name; the example near the end
        shows this clearly.

   TTL
        Standard DNS meaning.

   Class
        Standard DNS meaning.

   Priority
        As for MX, the priority of this target host. A client MUST
        attempt to contact the target host with the lowest-numbered
        priority it can reach; target hosts with the same priority
        SHOULD be tried in pseudorandom order. The range is 0-65535.

Weight
     Load balancing mechanism. When selecting a target host among
     the those that have the same priority, the chance of trying this
     one first SHOULD be proportional to its weight. The range of
     this number is 1-65535. Domain administrators are urged to use
     Weight 0 when there isn't any load balancing to do, to make the
     RR easier to read for humans (less noisy).

Port
     The port on this target host of this service. The range is
     0-65535. This is often as specified in Assigned Numbers but
     need not be.

Target
     As for MX, the domain name of the target host. There MUST be
     one or more A records for this name. Implementors are urged, but
     not required, to return the A record(s) in the Additional Data
     section. Name compression is to be used for this field.

     A Target of "." means that the service is decidedly not
     available at this domain.

Domain administrator advice

  Asking everyone to update their telnet (for example) clients when the
  first internet site adds a SRV RR for Telnet/TCP is futile (even if
  desirable). Therefore SRV will have to coexist with A record lookups
  for a long time, and DNS administrators should try to provide A
  records to support old clients:

    - Where the services for a single domain are spread over several
      hosts, it seems advisable to have a list of A RRs at the same
      DNS node as the SRV RR, listing reasonable (if perhaps
      suboptimal) fallback hosts for Telnet, NNTP and other protocols
      likely to be used with this name. Note that some programs only
      try the first address they get back from e.g. gethostbyname(),
      and we don't know how widespread this behaviour is.

    - Where one service is provided by several hosts, one can either
      provide A records for all the hosts (in which case the round-
      robin mechanism, where available, will share the load equally)
      or just for one (presumably the fastest).

    - If a host is intended to provide a service only when the main
      server(s) is/are down, it probably shouldn't be listed in A
      records.

- Hosts that are referenced by backup A records must use the port
  number specified in Assigned Numbers for the service.

Currently there's a practical limit of 512 bytes for DNS replies.
Until all resolvers can handle larger responses, domain
administrators are strongly advised to keep their SRV replies below
512 bytes.

All round numbers, wrote Dr. Johnson, are false, and these numbers
are very round: A reply packet has a 30-byte overhead plus the name
of the service ("telnet.tcp.asdf.com" for instance); each SRV RR adds
20 bytes plus the name of the target host; each NS RR in the NS
section is 15 bytes plus the name of the name server host; and
finally each A RR in the additional data section is 20 bytes or so,
and there are A's for each SRV and NS RR mentioned in the answer.
This size estimate is extremely crude, but shouldn't underestimate
the actual answer size by much. If an answer may be close to the
limit, using e.g. "dig" to look at the actual answer is a good idea.

The "Weight" field

Weight, the load balancing field, is not quite satisfactory, but the
actual load on typical servers changes much too quickly to be kept
around in DNS caches. It seems to the authors that offering
administrators a way to say "this machine is three times as fast as
that one" is the best that can practically be done.

The only way the authors can see of getting a "better" load figure is
asking a separate server when the client selects a server and
contacts it. For short-lived services like SMTP an extra step in the
connection establishment seems too expensive, and for long-lived
services like telnet, the load figure may well be thrown off a minute
after the connection is established when someone else starts or
finishes a heavy job.

The Port number

Currently, the translation from service name to port number happens
at the client, often using a file such as /etc/services.

Moving this information to the DNS makes it less necessary to update
these files on every single computer of the net every time a new
service is added, and makes it possible to move standard services out
of the "root-only" port range on unix.

Usage rules

   A SRV-cognizant client SHOULD use this procedure to locate a list of
   servers and connect to the preferred one:

        Do a lookup for QNAME=service.protocol.target, QCLASS=IN,
        QTYPE=SRV.

        If the reply is NOERROR, ANCOUNT>0 and there is at least one SRV
        RR which specifies the requested Service and Protocol in the
        reply:

             If there is precisely one SRV RR, and its Target is "."
             (the root domain), abort.

             Else, for all such RR's, build a list of (Priority, Weight,
             Target) tuples

             Sort the list by priority (lowest number first)

             Create a new empty list

             For each distinct priority level
                  While there are still elements left at this priority
                  level
                       Select an element randomly, with probability
                       Weight, and move it to the tail of the new list

             For each element in the new list

                  query the DNS for A RR's for the Target or use any
                  RR's found in the Additional Data secion of the
                  earlier SRV query.

                  for each A RR found, try to connect to the (protocol,
                  address, service).

        else if the service desired is SMTP

             skip to RFC 974 (MX).

        else

             Do a lookup for QNAME=target, QCLASS=IN, QTYPE=A

             for each A RR found, try to connect to the (protocol,
             address, service)

   Notes:

      - Port numbers SHOULD NOT be used in place of the symbolic service
        or protocol names (for the same reason why variant names cannot
        be allowed: Applications would have to do two or more lookups).

      - If a truncated response comes back from an SRV query, and the
        Additional Data section has at least one complete RR in it, the
        answer MUST be considered complete and the client resolver
        SHOULD NOT retry the query using TCP, but use normal UDP queries
        for A RR's missing from the Additional Data section.

      - A client MAY use means other than Weight to choose among target
        hosts with equal Priority.

      - A client MUST parse all of the RR's in the reply.

      - If the Additional Data section doesn't contain A RR's for all
        the SRV RR's and the client may want to connect to the target
        host(s) involved, the client MUST look up the A RR(s). (This
        happens quite often when the A RR has shorter TTL than the SRV
        or NS RR's.)

      - A future standard could specify that a SRV RR whose Protocol was
        TCP and whose Service was SMTP would override RFC 974's rules
        with regard to the use of an MX RR. This would allow firewalled
        organizations with several SMTP relays to control the load
        distribution using the Weight field.

      - Future protocols could be designed to use SRV RR lookups as the
        means by which clients locate their servers.

Fictional example

   This is (part of) the zone file for asdf.com, a still-unused domain:

        $ORIGIN asdf.com.
        @   SOA server.asdf.com. root.asdf.com. (
                            1995032001 3600 3600 604800 86400)
                        NS server.asdf.com.
                        NS ns1.ip-provider.net.
                        NS ns2.ip-provider.net.
        ftp.tcp   SRV 0 0 21 server.asdf.com.
        finger.tcp   SRV 0 0 79 server.asdf.com.
        ; telnet - use old-slow-box or new-fast-box if either is
        ; available, make three quarters of the logins go to
        ; new-fast-box.
        telnet.tcp   SRV 0 1 23 old-slow-box.asdf.com.

```
                         SRV 0 3 23 new-fast-box.asdf.com.
        ; if neither old-slow-box or new-fast-box is up, switch to
        ; using the sysdmin's box and the server
                         SRV 1 0 23 sysadmins-box.asdf.com.
                         SRV 1 0 23 server.asdf.com.
        ; HTTP - server is the main server, new-fast-box is the backup
        ; (On new-fast-box, the HTTP daemon runs on port 8000)
        http.tcp         SRV 0 0 80 server.asdf.com.
                         SRV 10 0 8000 new-fast-box.asdf.com.
        ; since we want to support both http://asdf.com/ and
        ; http://www.asdf.com/ we need the next two RRs as well
        http.tcp.www     SRV 0 0 80 server.asdf.com.
                         SRV 10 0 8000 new-fast-box.asdf.com.
        ; SMTP - mail goes to the server, and to the IP provider if
        ; the net is down
        smtp.tcp         SRV 0 0 25 server.asdf.com.
                         SRV 1 0 25 mailhost.ip-provider.net.
        @                MX  0 server.asdf.com.
                         MX  1 mailhost.ip-provider.net.
        ; NNTP - use the IP providers's NNTP server
        nntp.tcp         SRV 0 0 119 nntphost.ip-provider.net.
        ; IDB is an locally defined protocol
        idb.tcp          SRV  0 0 2025 new-fast-box.asdf.com.
        ; addresses
        server           A   172.30.79.10
        old-slow-box     A   172.30.79.11
        sysadmins-box    A   172.30.79.12
        new-fast-box     A   172.30.79.13
        ; backup A records - new-fast-box and old-slow-box are
        ; included, naturally, and server is too, but might go
        ; if the load got too bad
        @                A   172.30.79.10
                         A   172.30.79.11
                         A   172.30.79.13
        ; backup A RR for www.asdf.com
        www              A       172.30.79.10
        ; NO other services are supported
        *.tcp            SRV  0 0 0 .
        *.udp            SRV  0 0 0 .
```

   In this example, a telnet connection to "asdf.com." needs an SRV
   lookup of "telnet.tcp.asdf.com." and possibly A lookups of "new-
   fast-box.asdf.com." and/or the other hosts named. The size of the
   SRV reply is approximately 365 bytes:

      30 bytes general overhead
      20 bytes for the query string, "telnet.tcp.asdf.com."
      130 bytes for 4 SRV RR's, 20 bytes each plus the lengths of "new-

         fast-box", "old-slow-box", "server" and "sysadmins-box" -
         "asdf.com" in the query section is quoted here and doesn't
         need to be counted again.
      75 bytes for 3 NS RRs, 15 bytes each plus the lengths of
         "server", "ns1.ip-provider.net." and "ns2" - again, "ip-
         provider.net." is quoted and only needs to be counted once.
      120 bytes for the 6 A RR's mentioned by the SRV and NS RR's.

Refererences

   RFC 1918: Rekhter, Y., Moskowitz, R., Karrenberg, D., de Groot, G.,
         and E. Lear, "Address Allocation for Private Internets",
         RFC 1918, February 1996.

   RFC 1916 Berkowitz, H., Ferguson, P, Leland, W. and P. Nesser,
         "Enterprise Renumbering: Experience and Information
         Solicitation", RFC 1916, February 1996.

   RFC 1912 Barr, D., "Common DNS Operational and Configuration
         Errors", RFC 1912, February 1996.

   RFC 1900: Carpenter, B., and Y. Rekhter, "Renumbering Needs Work",
         RFC 1900, February 1996.

   RFC 1920: Postel, J., "INTERNET OFFICIAL PROTOCOL STANDARDS",
         STD 1, RFC 1920, March 1996.

   RFC 1814: Gerich, E., "Unique Addresses are Good", RFC 1814, June
            1995.

   RFC 1794: Brisco, T., "DNS Support for Load Balancing", April 1995.

   RFC 1713: Romao, A., "Tools for DNS debugging", November 1994.

   RFC 1712: Farrell, C., Schulze, M., Pleitner, S., and D. Baldoni,
         "DNS Encoding of Geographical Location", RFC 1712, November
         1994.

   RFC 1706: Manning, B. and R. Colella, "DNS NSAP Resource Records",
         RFC 1706, October 1994.

   RFC 1700: Reynolds, J., and J. Postel, "ASSIGNED NUMBERS",
         STD 2, RFC 1700, October 1994.

   RFC 1183: Ullmann, R., Mockapetris, P., Mamakos, L., and
         C. Everhart, "New DNS RR Definitions", RFC 1183, November
         1990.

    RFC 1101: Mockapetris, P., "DNS encoding of network names and other
          types", RFC 1101, April 1989.

    RFC 1035: Mockapetris, P., "Domain names - implementation and
          specification", STD 13, RFC 1035, November 1987.

    RFC 1034: Mockapetris, P., "Domain names - concepts and
          facilities", STD 13, RFC 1034, November 1987.

    RFC 1033: Lottor, M., "Domain administrators operations guide",
          RFC 1033, November 1987.

    RFC 1032: Stahl, M., "Domain administrators guide", RFC 1032,
          November 1987.

    RFC 974: Partridge, C., "Mail routing and the domain system",
          STD 14, RFC 974, January 1986.

Security Considerations

    The authors believes this RR to not cause any new security problems.
    Some problems become more visible, though.

      - The ability to specify ports on a fine-grained basis obviously
        changes how a router can filter packets. It becomes impossible
        to block internal clients from accessing specific external
        services, slightly harder to block internal users from running
        unautorised services, and more important for the router
        operations and DNS operations personnel to cooperate.

      - There is no way a site can keep its hosts from being referenced
        as servers (as, indeed, some sites become unwilling secondary
        MXes today). This could lead to denial of service.

      - With SRV, DNS spoofers can supply false port numbers, as well as
        host names and addresses. The authors do not see any practical
        effect of this.

    We assume that as the DNS-security people invent new features, DNS
    servers will return the relevant RRs in the Additional Data section
    when answering an SRV query.

Authors' Addresses

    Arnt Gulbrandsen
    Troll Tech
    Postboks 6133 Etterstad
    N-0602 Oslo
    Norway

    Phone: +47 22646966
    EMail: agulbra@troll.no

    Paul Vixie
    Vixie Enterprises
    Star Route 159A
    Woodside, CA 94062

    Phone: (415) 747-0204
    EMail: paul@vix.com