   Using DNS SRV to Specify a Global File Namespace with NFS Version 4

Abstract

   The NFS version 4 (NFSv4) protocol provides a mechanism for a
   collection of NFS file servers to collaborate in providing an
   organization-wide file namespace. The DNS SRV Resource Record (RR)
   allows a simple way for an organization to publish the root of its
   file system namespace, even to clients that might not be intimately
   associated with such an organization. The DNS SRV RR can be used to
   join these organization-wide file namespaces together to allow
   construction of a global, uniform NFS file namespace.

Status of This Memo

   This is an Internet Standards Track document.

   This document is a product of the Internet Engineering Task Force
   (IETF). It represents the consensus of the IETF community. It has
   received public review and has been approved for publication by the
   Internet Engineering Steering Group (IESG). Further information on
   Internet Standards is available in Section 2 of RFC 5741.

   Information about the current status of this document, any errata,
   and how to provide feedback on it may be obtained at
   http://www.rfc-editor.org/info/rfc6641.

Copyright Notice

Table of Contents

1. Background

   Version 4 of the NFS protocol [RFC3530] introduced the fs_locations
   attribute. Use of this attribute was elaborated further in the NFSv4
   minor version 1 protocol [RFC5661], which also defined an extended
   version of the attribute as fs_locations_info. With the advent of
   these attributes, NFS servers can cooperate to build a file namespace
   that crosses server boundaries. The fs_locations and
   fs_locations_info attributes are used as referrals, so that a file
   server may indicate to its client that the file name tree beneath a
   given name in the server is not present on itself but is represented
   by a file system in some other set of servers. The mechanism is
   general, allowing servers to describe any file system as being
   reachable by requests to any of a set of servers. Thus, starting
   with a single NFSv4 server, using these referrals, an NFSv4 client
   could see a large namespace associated with a collection of
   interrelated NFSv4 file servers. An organization could use this
   capability to construct a uniform file namespace for itself.

   An organization might wish to publish the starting point for this
   namespace to its clients. In many cases, the organization will want
   to publish this starting point to a broader set of possible clients.
   At the same time, it is useful to require that clients know only the
   smallest amount of information in order to locate the appropriate
   namespace. Also, that required information should be constant
   through the life of an organization if the clients are not to require
   reconfiguration as administrative events change, for instance, a
   server's name or address.

2. Requirements Notation

   The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
   "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this
   document are to be interpreted as described in [RFC2119].

3. Use of the SRV Resource Record in DNS

   Providing an organization's published file system namespace is a
   service, and the DNS [RFC1034][RFC1035] provides methods for
   discovery of that service. This standard defines a mapping from a
   DNS name to the NFS file system(s) providing the root of the file
   system namespace associated with that DNS name; such file systems are
   called "domain root" file systems. From such file systems, like
   other NFS file systems, an NFS client can use the standard NFS
   mechanisms to navigate the rest of the NFS file servers that make up
   the file system namespace for the given domain.

   Such domain root file systems are mounted at a conventional point in
   the NFS client namespace. The mechanism results in a uniform cross-
   organizational file namespace, similar to that seen in both AFS
   [AFS][RFC5864] and Distributed Computing Environment / Distributed
   File System (DCE/DFS) [DFS]. An NFS client need know only the domain
   name for an organization in order to locate the file namespace
   published by that organization.

   The DNS SRV RR type [RFC2782] is used to locate domain root file
   servers. The format of the DNS SRV record is as follows:

      _Service._Proto.Name TTL Class SRV Priority Weight Port Target

   The Service name used is "_nfs-domainroot", in conformance with RFC
   6335 [RFC6335]. The Protocol name used is "_tcp", for NFS service
   over TCP. Future NFS services using other protocols MUST use another
   protocol name. The "_udp" label MUST NOT be used to imply use of UDP
   with NFSv4, as neither RFC 3530 [RFC3530] nor RFC 5661 [RFC5661]
   defines NFSv4 over UDP. The Target fields give the domain names of
   the NFS servers that export file systems for the domain's root. An
   NFS client may then interpret any of the exported root file systems
   as the root of the file system published by the organization with the
   given domain name.

   The domain root service is not useful for NFS versions prior to
   version 4, as the fs_locations attribute was introduced only in NFSv4
   (as described in Section 1).

   In order to allow the NFSv4 servers as given to export a variety of
   file systems, those file servers MUST export the given domain's root
   file systems at "/.domainroot/{Name}" within their pseudo-file
   systems, where the "{Name}" is the name of the organization as used
   in the SRV RR.

   As an example, suppose a client wished to locate the root of the file
   system published by organization example.net. The DNS servers for
   the domain would publish records like

      $ORIGIN example.net.
      _nfs-domainroot._tcp IN SRV 0 0 2049 nfs1tr.example.net.
      _nfs-domainroot._tcp IN SRV 1 0 18204 nfs2ex.example.net.

The resulting domain names nfs1tr.example.net and nfs2ex.example.net
indicate NFSv4 file servers that export the root of the published
namespace for the example.net domain. In accordance with RFC 2782
[RFC2782], these records are to be interpreted using the Priority and
Weight field values, selecting an appropriate file server with which
to begin a network conversation. The two file servers would export
file systems that would be found at "/.domainroot/example.net" in
their pseudo-file systems, which clients would mount. Clients then
carry out subsequent accesses in accordance with the ordinary NFSv4
protocol. The first record uses the port number 2049 assigned to
NFS, and another port is specified for the second record; the NFS
servers would provide NFS service at their indicated port numbers,
and NFS clients would connect to the service via the corresponding
port numbers on those indicated servers.

Other file system protocols could make use of the same domain root
abstraction, but it is necessary to use different Service names not
specified here.

4. Integration with Use of NFS Version 4

NFSv4 clients adhering to this specification implement a special
directory, analogous to an Automounter [AMD1][AMD2] directory, the
entries in which are domain names that have recently been traversed.
When an application attempts to traverse a new name in that special
directory, the NFSv4 client consults DNS to obtain the SRV data for
the given name, and if successful, it mounts the indicated file
system(s) in that name in the special directory. The goal is that
NFSv4 applications will be able to look up an organization's domain
name in the special directory, and the NFSv4 client will be able to
discover the file system that the organization publishes. Entries in
the special directory will be domain names, and they will each appear
to the application as a directory name pointing to the root directory
of the file system published by the organization responsible for that
domain name.

As noted in Section 3, the domain root service is not useful for NFS
versions prior to version 4.

4.1. Globally Useful Names: Conventional Mount Point

In order for the inter-organizational namespace to function as a
global file namespace, the client-side mount point for that namespace
must be the same on different clients. Conventionally, on Portable
Operating System Interface (POSIX) machines, the name "/nfs4/" is
used so that names on one machine will be directly usable on any
machine. Thus, the example.net published file system would be
accessible as

          /nfs4/example.net/

   on any POSIX client. Using this convention, "/nfs4/" is the name of
   the special directory that is populated with domain names, leading to
   file servers and file systems that capture the results of SRV record
   lookups.

4.2. Mount Options

   SRV records are necessarily less complete than the information in the
   existing NFSv4 attributes fs_locations [RFC3530] or fs_locations_info
   [RFC5661]. For the rootpath field of fs_location, or the fli_fs_root
   field of fs_locations_info, NFS servers MUST use the "/.domainroot/
   {Name}" string. Thus, the servers listed as targets for the SRV RRs
   MUST export the root of the organization's published file system as
   the directory "/.domainroot/{Name}" (for the given organization Name)
   in their exported NFS namespaces. For example, for organization
   example.net, the directory "/.domainroot/example.net" would be used.

   Section 11 of the NFSv4.1 document [RFC5661] describes the approach
   that an NFS client should take to navigate fs_locations_info
   information.

   The process of mounting an organization's namespace should permit the
   use of what is likely to impose the lowest cost on the server. Thus,
   the NFS client SHOULD NOT insist on using a writable copy of the file
   system if read-only copies exist, or a zero-age copy rather than a
   copy that may be a little older. The organization's file system
   representatives can be navigated to provide access to higher-cost
   properties such as writability or freshness as necessary, but the
   default use when navigating to the base information for an
   organization ought to be as low-overhead as possible.

4.3. File System Integration Issues

   The result of the DNS search SHOULD appear as a (pseudo-)directory in
   the client namespace. A further refinement is RECOMMENDED: that only
   fully qualified domain names appear as directories. That is, in many
   environments, DNS names may be abbreviated from their fully qualified
   form. In such circumstances, multiple names might be given to NFS
   clients that all resolve to the same DNS SRV RRs. The abbreviated
   form SHOULD be represented in the client's namespace cache as a
   symbolic link, pointing to the fully qualified name. This will allow
   pathnames obtained with, say, getcwd() to include the DNS name that
   is most likely to be usable outside the scope of any particular DNS
   abbreviation convention.

4.4. Multicast DNS

   Location of the NFS domain root by this SRV record is intended to be
   performed with unicast by using the ordinary DNS [RFC1034][RFC1035]
   protocol.

   This document does not define the use of this DNS SRV record format
   in conjunction with Multicast DNS (mDNS). While mDNS could be used
   to locate a local domain root via these SRV records, no other
   domain's root could be discovered. This means that mDNS has too
   little value to use in locating NFSv4 domain roots.

5. Where Is This Integration Carried Out?

   The NFS client is responsible for interpreting SRV records. Using
   something like Automounter [AMD1] [AMD2] technology, the client
   interprets names under a particular directory, by first discovering
   the appropriate file system to mount and then mounting it in the
   specified place in the client namespace before returning control to
   the application doing a lookup. The result of the DNS lookup should
   be cached (obeying Time to Live (TTL)) so that the result could be
   returned more quickly the next time.

6. Security Considerations

   This functionality introduces a new reliance of NFSv4 on the
   integrity of DNS. Forged SRV records in DNS could cause the NFSv4
   client to connect to the file servers of an attacker, rather than the
   legitimate file servers of an organization. This is similar to
   attacks that can be made on the base NFSv4 protocol, if server names
   are given in fs_location attributes: the client can be made to
   connect to the file servers of an attacker, not the file servers
   intended to be the target for the fs_location attributes.

   If DNS Security Extensions (DNSSEC) [RFC4033] is available, it SHOULD
   be used to avoid both such attacks. Domain-based service principal
   names are an additional mechanism that also apply in this case, and
   it would be prudent to use them. They provide a mapping from the
   domain name that the user specified to names of security principals
   used on the NFSv4 servers that are indicated as the targets in the
   SRV records (as providing file service for the root file systems).

   With domain-based service principal names, the idea is that one wants
   to authenticate {nfs, domainname, host.fqdn}, not simply {nfs,
   host.fqdn}, when the server is a domain's root file server obtained
   through a DNS SRV RR lookup that may or may not have been secure.

The domain administrator can thus ensure that only domain root NFSv4
servers have credentials for such domain-based service principal
names.

Domain-based service principal names are defined in RFCs 5178
[RFC5178] and 5179 [RFC5179]. To make use of RFC 5178's domain-based
names, the syntax for "domain-based-name" MUST be used with a service
of "nfs", a domain matching the name of the organization whose root
file system is being sought, and a hostname given in the target of
the DNS SRV RR. Thus, in the example above, two file servers
(nfs1tr.example.net and nfs2ex.example.net) are located as hosting
the root file system for the organization example.net. To
communicate with, for instance, the second of the given file servers,
Generic Security Service Application Program Interface (GSS-API) is
used with the name-type of GSS_C_NT_DOMAINBASED_SERVICE defined in
RFC 5178 and with a symbolic name of

        nfs@example.net@nfs2ex.example.net

in order to verify that the named server (nfs2ex.example.net) is
authorized to provide the root file system for the example.net
organization.

NFSv4 itself contains a facility for the negotiation of security
mechanisms to be used between NFS clients and NFS servers. Section
3.3 of RFC 3530 [RFC3530] and Section 2.6 of RFC 5661 [RFC5661] both
describe how security mechanisms are to be negotiated. As such,
there is no need for this document to describe how that negotiation
is to be carried out when the NFS client contacts the NFS server for
the specified domain root file system(s).

Using SRV records to advertise the locations of NFS servers may
expose those NFS servers to attacks. Organizations should carefully
consider whether they wish their DNS servers to respond
differentially to different DNS clients, perhaps exposing their SRV
records to only those DNS requests that originate within a given
perimeter, in order to reduce this exposure.

7. IANA Considerations

   IANA has assigned a new Service name without an associated port
   number (as defined in RFC 6335 [RFC6335]) for TCP. For this new
   Service, the Reference is this document.

   Service name: nfs-domainroot
   Transport Protocol(s) TCP
   Assignee (REQUIRED) IESG (iesg@ietf.org)
   Contact (REQUIRED)   IETF Chair (chair@ietf.org)
   Description (REQUIRED) NFS service for the domain root, the root of
                      an organization's published file namespace.
   Reference (REQUIRED) This document
   Port Number (OPTIONAL)
   Service Code (REQUIRED for DCCP only)
   Known Unauthorized Uses (OPTIONAL)
   Assignment Notes (OPTIONAL)

8. References

8.1. Normative References

   [RFC1034]   Mockapetris, P., "Domain names - concepts and facilities",
               STD 13, RFC 1034, November 1987.

   [RFC1035]   Mockapetris, P., "Domain names - implementation and
               specification", STD 13, RFC 1035, November 1987.

   [RFC2119]   Bradner, S., "Key words for use in RFCs to Indicate
               Requirement Levels", BCP 14, RFC 2119, March 1997.

   [RFC2782]   Gulbrandsen, A., Vixie, P., and L. Esibov, "A DNS RR for
               specifying the location of services (DNS SRV)", RFC 2782,
               February 2000.

   [RFC3530]   Shepler, S., Callaghan, B., Robinson, D., Thurlow, R.,
               Beame, C., Eisler, M., and D. Noveck, "Network File System
               (NFS) version 4 Protocol", RFC 3530, April 2003.

   [RFC4033]   Arends, R., Austein, R., Larson, M., Massey, D., and S.
               Rose, "DNS Security Introduction and Requirements",
               RFC 4033, March 2005.

   [RFC5178]   Williams, N. and A. Melnikov, "Generic Security Service
               Application Program Interface (GSS-API)
               Internationalization and Domain-Based Service Names and
               Name Type", RFC 5178, May 2008.

   [RFC5179] Williams, N., "Generic Security Service Application
              Program Interface (GSS-API) Domain-Based Service Names
              Mapping for the Kerberos V GSS Mechanism", RFC 5179,
              May 2008.

   [RFC5661] Shepler, S., Ed., Eisler, M., Ed., and D. Noveck, Ed.,
              "Network File System (NFS) Version 4 Minor Version 1
              Protocol", RFC 5661, January 2010.

   [RFC5864] Allbery, R., "DNS SRV Resource Records for AFS", RFC 5864,
              April 2010.

   [RFC6335] Cotton, M., Eggert, L., Touch, J., Westerlund, M., and S.
              Cheshire, "Internet Assigned Numbers Authority (IANA)
              Procedures for the Management of the Service Name and
              Transport Protocol Port Number Registry", BCP 165,
              RFC 6335, August 2011.

8.2. Informative References

   [AFS]   Howard, J., "An Overview of the Andrew File System", Proc.
              USENIX Winter Tech. Conf. Dallas, February 1988.

   [AMD1]   Pendry, J. and N. Williams, "Amd: The 4.4 BSD Automounter
              Reference Manual", March 1991,
              <http://docs.freebsd.org/info/amdref/amdref.pdf>.

   [AMD2]   Crosby, M., "AMD--AutoMount Daemon", Linux Journal,
              35es Article 4, March 1997.

   [DFS]   Kazar, M., Leverett, B., Anderson, O., Apostolides, V.,
              Bottos, B., Chutani, S., Everhart, C., Mason, W., Tu, S.,
              and E. Zayas, "DEcorum File System Architectural
              Overview", Proc. USENIX Summer Conf. Anaheim, Calif.,
              June 1990.

Authors' Addresses

   Craig Everhart
   NetApp
   800 Cranberry Woods Drive, Ste. 300
   Cranberry Township, PA 16066
   USA

   Phone: +1 724 741 5101
   EMail: everhart@netapp.com

   W.A. (Andy) Adamson
   NetApp
   495 East Java Drive
   Sunnyvale, CA 94089
   USA

   Phone: +1 734 665 1204
   EMail: andros@netapp.com

   Jiaying Zhang
   Google
   604 Arizona Avenue
   Santa Monica, CA 90401
   USA

   Phone: +1 310 309 6884
   EMail: jiayingz@google.com