Network Working Group	C. Everhart
Internet-Draft	NetApp
Intended status: Standards Track	W.A. Adamson
Expires: April 09, 2012	NetApp
	J. Zhang
	Google
	October 07, 2011

Using DNS SRV to Specify a Global File Name Space with NFS version 4
draft-ietf-nfsv4-federated-fs-dns-srv-namespace-09.txt

Abstract

The NFS version 4 protocol provides a mechanism for a collection of NFS file servers to collaborate in providing an organization-wide file name space. The DNS SRV RR allows a simple and appropriate way for an organization to publish the root of its filesystem name space, 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 name spaces together to allow construction of a global, uniform NFS file name space.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.
Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet- Drafts is at http://datatracker.ietf.org/drafts/current/.
Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."
This Internet-Draft will expire on April 09, 2012.

Table of Contents

1. 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].

2. Background

Version 4 of the NFS protocol [RFC3530] introduced the fs_locations attribute. Use of this attribute was elaborated further in the NFS Version 4 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 name space 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 filesystem in some other set of servers. The mechanism is general, allowing servers to describe any filesystem as being reachable by requests to any of a set of servers. Thus, starting with a single NFS Version 4 server, using these referrals, an NFS Version 4 client could see a large name space associated with a collection of interrelated NFS Version 4 file servers. An organization could use this capability to construct a uniform file name space for itself.

An organization might wish to publish the starting point for this name space 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 clients to know only the smallest amount of information in order to locate the appropriate name space. Simultaneously, 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.

3. Use of SRV Resource Record in DNS

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

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

The DNS SRV resource record 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 is "_nfs._domainroot". The Protocol as of this writing could be either "_tcp" or "_sctp"; version 4 NFS is not defined over UDP. Other transport protocols could be defined in the future, and SRV records that advertise domain root file services with other transport protocols would use the same Service name. The Target fields give the domain names of the NFS servers that export filesystems for the domain's root. An NFS client may then interpret any of the exported root filesystems as the filesystem published by the organization with the given domain name.

In order to allow the NFSv4 servers so given to export a variety of filesystems, those file servers MUST export the given domain's root filesystems at "/.domainroot-{Name}" within their pseudo-filesystems, 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 filesystem 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 2049 nfs2ex.example.net.
```

The result domain names nfs1tr.example.net and nfs2ex.example.net indicate NFS Version 4 file servers that export the root of the published name space 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 filesystems that would be found at "/.domainroot-example.net" in their pseudo-filesystems, which clients would mount. Clients then carry out subsequent accesses in accordance with the ordinary NFS Version 4 protocol.

Other filesystem protocols could make use of the same "domain root" abstraction, but this document discusses only the SRV records indicating NFS servers.

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 filesystem(s) in that name in the special directory. The goal is that NFS applications will be able to lookup an organization's domain name in the special directory, and the NFSv4 client will be able to discover the filesystem that that 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 filesystem published by the organization responsible for that domain name.
This DNS SRV record evaluation could, in principle, be done either in the NFSv4 client or in an NFSv4 server. In either case, the result would appear the same to applications on the NFSv4 client. This is discussed further in section 5 of this document.

4.1. Globally-useful names: conventional mount point

In order that the inter-organizational name space function as a global name space, the client-side mount point for that name space must be the same on different clients. Conventionally, on POSIX machines, the name /nfs4/ is be used so that names on one machine will be directly usable on any machine. Thus, the example.net published filesystem 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 filesystems that capture the results of SRV record lookups.

4.2. Mount options

SRV records are necessarily less complete than the information in the existing NFS Version 4 attributes fs_locations [RFC3530] or fs_locations_info [RFC5661]. For the rootpath field of fs_location, or the fli_fs_root of fs_locations_info, NFS servers MUST use the "/.domainroot-{Name}" string. Thus, the servers listed as targets for the SRV resource records MUST export the root of the organization's published filesystem 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.
Chapter 11 of the NFS Version 4.1 document [RFC5661] describes the approach that an NFS client should take to navigating fs_locations_info information.
The process of mounting an organization's name space 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 filesystem if read-only copies exist, or a zero-age copy rather than a copy that may be a little older. We presume that the organization's file name space can be navigated to provide access to higher-cost properties such as writability or freshness as necessary, but that the default use when navigating to the base information for an organization ought to be as low-overhead as possible.
Because of the possible distinction between read-only and read-write versions of a filesystem, organizations MAY also publish the location of a writable instance of their domain root filesystems, and that NFSv4

clients would conventionally mount that filesystem under the organizational domain name preceded by a period ("."). Therefore, when names beginning with a period are looked up under the NFSv4 client's special directory, the SRV RR looked up in DNS uses a Service name of "_nfs._write._domainroot", and any indicated server (or servers) MUST export the writable instance at "/.domainroot-write-{Name}" for a domain name Name.

Extending the opening example, suppose a client wished to locate the read-only and read-write roots of the filesystem published by organization example.net. Suppose a read-write instance were hosted on server nfs1tr.example.net, and read-only instances were on that server and also on server nfs2ex.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 2049 nfs2ex.example.net.
_nfs._write._domainroot._tcp IN SRV 0 0 2049 nfs1tr.example.net.
```

The nfs1tr.example.net server would export filesystems at both "/.domainroot-example.net" (the read-only instance) and "/.domainroot-write-example.net" (the read-write instance). The nfs2ex.example.net server need export only the "/.domainroot-example.net" name for its read-only instance.

The read-write version of the filesystem would be mounted (upon use) under ".example.net" in the special directory, and a read-only version would be mounted under "example.net". Thus,

```
/nfs4/example.net/users

/nfs4/.example.net/users
```

would be a writable form of that same directory.

4.3. Filesystem integration issues

The result of the DNS search SHOULD appear as a (pseudo-)directory in the client name space. 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 name space cache as a symbolic link, pointing to the fully-qualified name, case-folded if the client filesystem is case-sensitive. 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, mDNS, and DNS-SD

Location of the NFS domain root does not involve IP-layer broadcast, multicast, or anycast communication.
It is not expected that this DNS SRV record format will be used in conjunction with Multicast DNS (mDNS) or DNS Service Discovery (DNS-SD). 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 with DNS-SD 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, discovering the appropriate filesystem to mount, and mounting it in the specified place in the client name space before returning control to the application doing a lookup. The result of the DNS lookup should be cached (obeying 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, not the 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 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 filesystems).
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 filesystem is being sought, and a hostname given in the target of the DNS SRV

resource record. Thus, in the example above, two file servers (nfs1tr.example.net and nfs2ex.example.net) are located as hosting the root filesystem for the organization example.net. To communicate with, for instance, the second of the given file servers, 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 filesystem 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 filesystem(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

This document requests the assignment of two new Service names without associated port numbers, each for both TCP and SCTP. For both Services, the Reference is this document.

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


Service name: _nfs._write._domainroot
Transport Protocol(s) TCP, SCTP
Assignee (REQUIRED)  IESG (iesg@ietf.org)
Contact (REQUIRED)   IETF Chair (chair@ietf.org)
Description (REQUIRED)  Writable file server for the NFS file service
                  for the domain root, the root of the organization's
                  published file name space
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

[RFC2119]	Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", March 1997.
[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.
[RFC5661]	Shepler, S., Eisler, M. and D. Noveck, Editors, "Network File System (NFS) Version 4 Minor Version 1 Protocol", RFC 5661, January 2010.
[RFC1034]	Mockapetris, P.V., "Domain Names - Concepts and Facilities", RFC 1034, November 1987.
[RFC1035]	Mockapetris, P.V., "Domain Names - Implementation and Specification", RFC 1035, November 1987.
[RFC2782]	Gulbrandsen, A., Vixie, P. and L. Esibov, "A DNS RR for specifying the location of services (DNS SRV)", RFC 2782, February 2000.
[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.
[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", RFC 6335, August 2011.

8.2. Informative References

[DFS]	Kazar, M.L., Leverett, B.W., Anderson, O.T., Apostolides, V., Bottos, B.A., Chutani, S., Everhart, C.F., Mason, W.A., Tu, S.-T. and E.R. Zayas, "DEcorum File System Architectural Overview", Proc. USENIX Summer Conf. Anaheim, Calif., June 1990.
[AMD1]	Pendry, J.-S. and N. Williams, "Amd: The 4.4 BSD Automounter Reference Manual", March 1991.
[AMD2]	Crosby, M., "AMD--AutoMount Daemon", Linux Journal 1997, 35es Article 4, March 1997.
[AFS]	Howard, J.H., "An Overview of the Andrew File System"", Proc. USENIX Winter Tech. Conf. Dallas, February 1988.

Authors' Addresses

Craig Everhart Everhart NetApp 800 Cranberry Woods Drive, Ste. 300 Cranberry Township, PA 16066 US Phone: +1 724 741 5101 EMail: everhart@netapp.com

W.A. (Andy) Adamson Adamson NetApp 495 East Java Drive Sunnyvale, CA 94089 US Phone: +1 734 665 1204 EMail: andros@netapp.com

Jiaying Zhang Zhang Google 604 Arizona Avenue Santa Monica, CA 90401 US Phone: +1 310 309 6884 EMail: jiayingz@google.com