Network Working Group	C. Everhart
Internet-Draft	W. Adamson
Intended status: Standards Track	NetApp
Expires: November 16, 2009	J. Zhang
	Google
	May 15, 2009

TOC

Using DNS SRV to Specify a Global File Name Space with NFS version 4
draft-ietf-nfsv4-federated-fs-dns-srv-namespace-01.txt

Status of this Memo

Copyright Notice

Abstract

The NFS version 4 protocol provides a natural way 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 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 version 4 file name space.

Table of Contents

---

1. Requirements notation

TOC

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] (Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels," March 1997.).

---

2. Background

TOC

With the advent of fs_locations attributes in the NFS Version 4 protocol [RFC3530] (Shepler, S., Callaghan, B., Robinson, D., Thurlow, R., Beame, C., Eisler, M., and D. Noveck, "Network File System (NFS) version 4 Protocol," April 2003.), NFS servers can cooperate to build a file name space that crosses server boundaries, as detailed in the description of referrals in [NB0510] (Noveck, D. and R. Burnett, "Next Steps for NFSv4 Migration/Replication," October 2005.). With NFS Version 4 referrals, a file server may indicate to its client that the file system 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 might be able to 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. Proposed Use of SRV Resource Record in DNS

TOC

Providing an organization's published file system name space is a service, and it is appropriate to use the DNS [RFC1035] (Mockapetris, P., “Domain Names - Implementation and Specification,” November 1987.) to locate it. As with the AFSDB resource record type [RFC1183] (Everhart, C., Mamakos, L., Ullmann, R., and P. Mockapetris, “New DNS RR Definitions,” October 1990.), the client need only utter the (relatively) constant domain name for an organization in order to locate its file system name space service. Once a client uses the DNS to locate one or more servers for the root of the organization's name space, it can use the standard NFS Version 4 mechanisms to navigate the remainder of the NFS servers for that organization. The use of this proposed mechanism results in a useful cross-organizational name space, just as in AFS [AFS] (Howard, J., “An Overview of the Andrew File System",” February 1988.) and DCE/DFS [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,” June 1990.) before it. A client need know only the name of the organization in order to locate the file system name space published by that organization.

We propose the use of the DNS SRV resource record type [RFC2782] to fulfill this function. The format of the DNS SRV record is as follows:

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

In our case, we use a Service name of "nfs4" and a conventional Protocol of "_tcp". The Target fields give the domain names of the NFS Version 4 servers that export root filesystems. An NFS Version 4 client

SHOULD interpret any of the exported pseudo-root filesystems as the filesystem published by the organization with the given domain name. Suppose a client wished to locate the root of the file system published by organization example.net. The DNS servers for the domain could publish records like

```
_nfs4._tcp IN SRV 0 0 2049 nfs1tr.example.net
_nfs4._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, 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. Subsequent accesses are carried out in accordance with ordinary NFS Version 4 protocol.

3.1. Deployment of the Resource Record

TOC

As with any DNS resource, any server installation needs to concern itself with the likely loads and effects of the presence of the resource record. The answers to requests for RRs might differ depending on what the server can tell about the client. For example, some RRs might be returned only to those clients inside some network perimeter (to provide an intranet service) and requests from other clients might be denied. As the RR directs the clients to ask for service from a given set of servers, the administrator should ensure that the identified servers can handle the expected load. Fortunately, the definition of the DNS SRV resource record offers a mechanism to distribute the load to multiple servers within a priority ordering.

---

4. Integration with Use of NFS Version 4

TOC

There are at least two remaining questions: whether this DNS SRV record evaluation is done in the NFS server or client, and also how the domain names of the organizations are passed to client or server. A third question is how this might produce a uniform global file name space, and what prefix should be used for such file names.
This specification anticipates that these SRV records will most commonly be used to define the second directory level in an inter-organizational file name space. This directory will be populated with domain names pointing to the file systems published for use under those domain names. Thus, the root directory for the file system published by

example.net will effectively be mounted underneath the example.net name in a second-level directory.
In general, a domain name will appear to a client as a directory name pointing to the root directory of the file system published by the organization responsible for that domain name.

4.1. Globally-useful names: conventional mount point

TOC

For the inter-organizational name space to be a global name space, it is useful for its mount point in local systems to be uniform as well. The name /nfs4/ SHOULD be 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 client. Using this convention, "/nfs4/" is a mount for a special file system that is populated with the results of SRV record lookups.

4.2. Mount options

TOC

SRV records are necessarily less complete than the information in the existing NFS Version 4 attributes fs_locations and the proposed fs_locations_info. For the rootpath field of fs_location, we assume that the empty string is adequate. Thus, the servers listed as targets for the SRV resource records should export the root of the organization's published file system as the pseudo-root in its exported namespace.
As for the other attributes in fs_locations_info, the recommended approach is for a client to make its first possible contact with any of the referred-to servers, obtain the fs_locations_info structure from that server, and use the information from that obtained structure as the basis for its judgment of whether it would be better to use a different server representative from the set of servers for that filesystem.
We recommend, though, that 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, we recommend that the client 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 currency as necessary, but that the default use when navigating to the base

information for an organization ought to be as low-overhead as possible.
One extension of this rule that we might choose to inherit from AFS, though, is to give a special meaning to the domain name of an organization preceded by a period ("."). It might be reasonable to have names mounting the filesystem for a period-prefixed domain name (e.g., ".example.net") attempt to mount only a read-write instance of that organization's root filesystem, rather than permitting the use of read-only instances of that filesystem. Thus,

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

might be a directory in a read-only instance of the root filesystem of the organization "example.net", while

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

would be a writable form of that same directory. A small benefit of following this convention is that names with the period prefix are treated as "hidden" in many operating systems, so that the visible name remains the lowest-overhead name.

4.3. File system integration issues

TOC

The result of the DNS search SHOULD appear as a (pseudo-)directory in the client name space, cached for a time no longer than the RR's TTL. A further refinement is advisable, and SHOULD be deployed: 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 file system code 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-canonicalized when appropriate. 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.

---

5. Where is this integration carried out?

TOC

Another consideration is what agent should be responsible for interpreting the SRV records. It could be done just as well by the client or by the server, though we expect that most clients will include this function themselves. Using something like Automounter [AMD] (Pendry, J. and N. Williams, “Amd: The 4.4 BSD Automounter

Reference Manual," March 1991.) technology, the client would be responsible for interpreting names under a particular directory, discovering the appropriate filesystem to mount, and mounting it in the appropriate place in the client name space before returning control to the application doing a lookup. Alternatively, one could imagine the existence of an NFS version 4 server that awaited similar domain-name lookups, then consulted the DNS SRV records to determine the servers for the indicated published file system, and then returned that information via NFS Version 4 attributes as a referral in the way outlined by Noveck and Burnett [NB0510] (Noveck, D. and R. Burnett, "Next Steps for NFSv4 Migration/Replication," October 2005.). In either case, the result of the DNS lookup should be cached (obeying TTL) so that the result could be returned more quickly the next time.
We strongly suggest that this functionality be implemented by NFS clients. While we recognize that it would be possible to configure clients so that they relied on a specially-configured server to do their SRV lookups for them, we feel that such a requirement would impose unusual dependencies and vulnerabilities for the deployers of such clients.

---

6. Relationship to DNS NFS4ID RR

TOC

This DNS use has no obvious relationship to the NFS4ID RR. The NFS4ID RR is a mechanism to help clients and servers configure themselves with respect to the domain strings used in "who" strings in ACL entries and in owner and group names. The authentication/authorization domain string of a server need have no direct relationship to the name of the organization that is publishing a file name space of which this server's filesystems form a part. At the same time, it might be seen as straightforward or normal for such a server to refer to the ownership of most of its files using a domain string with an evident relationship to that NFS4ID-given domain name, but this document imposes no such requirement.

---

7. Security Considerations

TOC

Naive use of the DNS may effectively give clients published server referrals that are intrusive substitutes for the servers intended by domain administrators.
It may be possible to build a trust chain by using DNSSEC [RFC4033] (Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "DNS Security Introduction and Requirements," March 2005.) to implement this function on the client, or by implementing this function on an NFS Version 4 server that uses DNSSEC and maintaining a trust relationship

with that server. This trust chain also breaks if the SRV interpreter accepts responses from insecure DNS zones. Thus, it would likely be prudent also to use domain-based service principal names for the servers for the root filesystems as indicated as the targets of the SRV records. The idea here 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 an insecure DNS SRV RR lookup. 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 [RFC3530] (Shepler, S., Callaghan, B., Robinson, D., Thurlow, R., Beame, C., Eisler, M., and D. Noveck, “Network File System (NFS) version 4 Protocol,” April 2003.) and 5179 [RFC3530] (Shepler, S., Callaghan, B., Robinson, D., Thurlow, R., Beame, C., Eisler, M., and D. Noveck, “Network File System (NFS) version 4 Protocol,” April 2003.). 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 should be 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.

8. References

TOC

8.1. Normative References

TOC

[RFC1034]	Mockapetris, P., “Domain Names - Concepts and Facilities,” RFC 1034, November 1987 (TXT).
[RFC1035]	Mockapetris, P., “Domain Names - Implementation and Specification,” RFC 1035, November 1987 (TXT).
[RFC2119]	Bradner, S., “Key words for use in RFCs to Indicate Requirement Levels,” 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 (TXT).
[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 (TXT).
[RFC4033]	Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "DNS Security Introduction and Requirements," RFC 4033, March 2005 (TXT).
[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 (TXT).
[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 (TXT).

8.2. Informative References

TOC

[AFS]	Howard, J., "An Overview of the Andrew File System",” Proc. USENIX Winter Tech. Conf. Dallas, February 1988.
[AMD]	Pendry, J. and N. Williams, "Amd: The 4.4 BSD Automounter Reference Manual," March 1991.
[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.
[NB0510]	Noveck, D. and R. Burnett, "Next Steps for NFSv4 Migration/Replication," October 2005.
[RFC1183]	Everhart, C., Mamakos, L., Ullmann, R., and P. Mockapetris, "New DNS RR Definitions," RFC 1183, October 1990 (TXT).

Authors' Addresses

TOC

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

Email:	everhart@netapp.com
	Andy Adamson
	NetApp
	495 East Java Drive
	Sunnyvale, CA 94089
	US
Phone:	+1 734 665 1204
Email:	andros@netapp.com
	Jiaying Zhang
	Google
	604 Arizona Avenue
	Santa Monica, CA 90401
	US
Phone:	+1 310 309 6884
Email:	jiayingz@google.com