Network Working Group	R. Allbery	
Internet-Draft	Stanford University	
Updates: 1183 (if approved)	February 25, 2010	
Intended status: Standards Track		
Expires: August 29, 2010		

TOC

DNS SRV Resource Records for AFS
draft-allbery-afs-srv-records-05

Abstract

This document specifies how to use DNS (Domain Name Service) SRV RRs (Resource Records) to locate services for the AFS distributed file system and how the priority and weight values of the SRV RR should be interpreted in the server ranking system used by AFS. It updates RFC 1183 to deprecate use of the AFSDB RR to locate AFS cell database servers and provides guidance for backward compatibility.

Internet Draft Comments

Comments are solicited. Please include the AFS Standardization mailing list at afs3-standardization@openafs.org as a recipient of any comments.

Status of this Memo

Table of Contents

---

1. Overview and Rationale

TOC

AFS (a registered trademark of IBM Corporation) is a distributed file system (see [AFS1] (Howard, J., Kazar, M., Menees, S., Nichols, D., Satyanarayanan, M., Sidebotham, R., and M. West, “Scale and Performance in a Distributed File System,” February 1988.) and [AFS2] (Howard, J., “An Overview of the Andrew File System,” February 1988.)). Its most widely-used implementations are [OPENAFS] (IBM Corporation, et al., “OpenAFS Documentation,” April 2000.) and [ARLA] (Assar Westerlund, et al., “Arla,” May 2001.).
AFS is organized administratively into cells. Each AFS cell consists of one or more Volume Location Database (VLDB) servers, one or more Protection Service (PTS) servers, and one or more file servers and volume servers, plus possibly additional services not relevant to this document. Data stored in AFS is divided into collections of files

called volumes. An AFS protocol client, when accessing a file within a specific AFS cell, first contacts a VLDB server for that cell to determine the file server for the AFS volume in which that file is located, and then contacts that file server directly to access the file. A client may also need to contact a PTS server for that cell to register before accessing files in that cell or to query protection database information.

An AFS client therefore needs to determine, for a given AFS cell, the VLDB and possibly the PTS servers for that cell. (Traditionally, the VLDB and PTS servers are provided by the same host.) Once the client is in contact with the VLDB server, it locates file and volume servers through AFS protocol queries to the VLDB server. Originally, VLDB server information was configured separately on each client in a file called the CellServDB file. [RFC1183] (Everhart, C., Mamakos, L., Ullmann, R., and P. Mockapetris, “New DNS RR Definitions,” October 1990.) specified the DNS RR (Resource Record) AFSDB to locate VLDB servers for AFS.

Subsequent to [RFC1183] (Everhart, C., Mamakos, L., Ullmann, R., and P. Mockapetris, “New DNS RR Definitions,” October 1990.), a general DNS RR was defined by [RFC2782] (Gulbrandsen, A., Vixie, P., and L. Esibov, “A DNS RR for specifying the location of services (DNS SRV),” February 2000.) for service location for any service. This DNS SRV RR has several advantages over the AFSDB RR:

- AFSDB RRs do not support priority or ranking, leaving AFS cell administrators without a way to indicate which VLDB servers clients should prefer.

- AFSDB RRs do not include protocol or port information, implicitly assuming that all VLDB servers will be contacted over the standard port and the UDP protocol. Future changes to the AFS protocol may require separate VLDB server lists for UDP and TCP traffic, and some uses of AFS, such as providing VLDB service for multiple cells from the same systems, require use of different ports.

- Clients using AFSDB RRs must assume that VLDB and PTS services are provided by the same host, but it may be useful to separate VLDB servers from PTS servers.

- DNS SRV RRs are in widespread use, whereas AFSDB RRs are a little-known and little-supported corner of the DNS protocol.

For those reasons, it is desirable to move AFS service location from the AFSDB RR to DNS SRV RRs.

TOC

2. Scope

This document describes the format and use of DNS SRV RRs for AFS service location and deprecates the AFSDB RR. It also provides guidance for transition from the AFSDB RR to DNS SRV RRs and recommendations for backward compatibility.
Documentation of the AFS protocol, the exact purpose and use of the VLDB and PTS services, and other information about AFS are outside the scope of this document.
AFSDB RRs may also be used for locating servers for the Open Software Foundation's (OSF) Distributed Computing Environment (DCE) authenticated naming system, as described in [RFC1183] (Everhart, C., Mamakos, L., Ullmann, R., and P. Mockapetris, “New DNS RR Definitions,” October 1990.). Service location for DCE servers is outside the scope of this document and not modified by this specification.

---

3. 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.).

---

4. DNS SRV RRs for AFS

TOC

The label of a DNS SRV RR, as defined in [RFC2782] (Gulbrandsen, A., Vixie, P., and L. Esibov, “A DNS RR for specifying the location of services (DNS SRV),” February 2000.), is:

```
_<service>._<proto>.<name>
```

The following values for <service> advertise servers providing AFS services:

afs3-vlserver: servers providing AFS VLDB services.

afs3-prserver: servers providing AFS PTS services.

Other AFS services, such as file and volume management services, are located through the VLDB service and therefore do not use DNS SRV RRs. <proto> MUST be "udp" for the current AFS protocol, which uses Rx over UDP. Other values may be used for future revisions of the AFS protocol supporting other protocols, such as Rx over TCP.

<name> MUST be the AFS cell name for which the identified server provides AFS services. Clients MUST query DNS SRV RRs only for a <name> value exactly matching the AFS cell of interest. They MUST NOT remove leading components to search for more general DNS SRV RRs. The AFS cell "prod.example.com" and the AFS cell "example.com" are entirely different cells in the AFS protocol and VLDB servers for the latter cannot provide information for the former.

> NOTE: As with AFSDB RRs, this means that DNS SRV RRs can only be used to locate AFS services for cells whose naming matches the structure of the DNS. This is not a requirement of the AFS protocol, but sites creating new AFS cells SHOULD use names that follow the structure of the DNS and which result in DNS SRV RRs under their administrative control. This both permits use of DNS SRV RRs instead of client configuration and helps avoid naming conflicts between separate AFS cells.

DNS SRV RRs include a priority and a weight. As defined in the DNS SRV RR specification [RFC2782] (Gulbrandsen, A., Vixie, P., and L. Esibov, “A DNS RR for specifying the location of services (DNS SRV),” February 2000.), clients MUST attempt to contact the target host with the lowest-numbered priority they can reach. AFS clients which use a ranked algorithm to determine which server to contact MUST therefore assign a sufficiently distinct ranks to targets with different priorities such that targets with a higher-numbered priority are only contacted if all targets with a lower-numbered priority are inaccessible. See Section 4.1 (Interpretation as AFS Preference Ranks) for more information.

If there are multiple targets with an equal priority, the weight value of the DNS SRV RR SHOULD be used as input to a weighted algorithm for selecting servers. As specified by [RFC2782] (Gulbrandsen, A., Vixie, P., and L. Esibov, “A DNS RR for specifying the location of services (DNS SRV),” February 2000.), larger weights SHOULD be given a proportionately higher probability of being selected. In the presence of records containing weights greater than 0, records with weight 0 should have a very small chance of being selected. A weight of 0 SHOULD be used if all targets with that priority are weighted equally. AFS clients MAY take into account network performance and other protocol metrics along with SRV RR weights when selecting servers, thereby possibly selecting different servers than a system based purely on the SRV RRs would indicate. However, such information MUST NOT override the priority information in the SRV RR.

DNS SRV RRs, like all DNS RRs, have a time-to-live (TTL), after which the SRV record information is no longer valid (see [RFC1034] (Mockapetris, P., “Domain names - concepts and facilities,” November 1987.)). DNS RRs SHOULD be discarded after their TTL, and the DNS query repeated. This applies to DNS SRV RRs for AFS as to any other DNS RR. Any information derived from the DNS SRV RRs, such as preference ranks, MUST be discarded when the DNS SRV RR is expired.

Implementations are not required to re-run the weighted server selection algorithm for each call. They MAY instead reuse the results of the algorithm until the DNS SRV RRs expire. Clients could therefore use a specific server for the lifetime of the DNS SRV records, which may affect the load distribution properties that DNS SRV records provide. Server operators should account for this effect when setting the TTL of those records.
AFS clients MAY remember which targets are inaccessible by that client and ignore those targets when determining which server to contact first. Clients which do this SHOULD have a mechanism to retry targets which were previously inaccessible and reconsider them according to their current priority and weight if they become accessible again.

4.1. Interpretation as AFS Preference Ranks

TOC

Several AFS implementations use a ranking algorithm that assigns numbers representing a preference rank to each server when the client first contacts that AFS cell and then prefers the server with the lowest rank unless that server goes down. Clients using this algorithm SHOULD assign their ranks as follows:

1. Sort targets by priority and assign a base rank value to each target based on its priority. Each base rank value MUST be sufficiently different from the base rank assigned to any higher-numbered priority that higher-numbered targets will only be attempted if lower-numbered targets cannot be reached. It MUST, in other words, be farther from the base rank of any distinct priority than any possible automatic adjustment in the rank. When calculating base ranks, observe that the numeric value of a priority has no meaning. Only the ordering of distinct priority values between multiple SRV RR targets needs to be reflected in the base ranks.

2. For each group of targets with the same priority, follow the algorithm in [RFC2782] (Gulbrandsen, A., Vixie, P., and L. Esibov, “A DNS RR for specifying the location of services (DNS SRV),” February 2000.) to order those targets. Then, assign those targets ranks formed by incrementing the base weight for that priority such that the first selected target has the lowest rank, the second selected target has the next lowest rank, and so on.

After assignment of ranks, the AFS client MAY then adjust the ranks dynamically based on network performance and other protocol metrics, provided that such adjustments are sufficiently small compared to the difference between base ranks that they cannot cause servers with a

higher-numbered priority to be contacted instead of a server with a lower-numbered priority.
The TTL of the DNS SRV RRs MUST be honored by invalidating and regenerating the server preference ranks with new DNS information once that TTL has expired. However, accumulated network and protocol metrics may be retained and reapplied to the new rankings.
AFS server preference ranks are conventionally numbers between 1 and 65535. DNS SRV RR priorities are numbers between 0 and 65535. Implementations following this algorithm could therefore encounter problems assigning sufficiently distinct base rank values in exceptional cases of very large numbers of DNS SRV RR targets with different priorities. However, an AFS cell configuration with thousands of DNS SRV RR targets for an AFS VLDB or PTS service with meaningfully distinct priorities is highly improbable. AFS client implementations encountering a DNS SRV RR containing targets with more distinct priority values than can be correctly represented as base ranks SHOULD fall back to generating ranks based solely on priorities, ignoring other rank inputs, and disabling dynamic adjustment of ranks. Implementations MUST be able to assign distinct base ranks as described above for at least ten distinct priority values.

---

5. Use of AFSDB RRs

TOC

Since many AFS client implementations currently support AFSDB RRs but do not support DNS SRV RRs, AFS cells providing DNS SRV RRs SHOULD also provide AFSDB RRs. However, be aware that AFSDB RRs do not provide priority or weighting information; all servers listed in ASFDB RRs are treated as equal. AFSDB RRs also do not provide port information.
An AFS cell using DNS SRV RRs SHOULD therefore also provide an AFSDB RR listing all AFS servers for which the following statements are all true:

- The server provides both VLDB and PTS service on the standard ports (7003 and 7002) respectively.
- The server provides these services via Rx over UDP.
- The server either has the lowest-numbered priority of those listed in the DNS SRV RRs or the AFS cell administrator believes it reasonable for clients using AFSDB RRs to use this server by preference.

The above is a default recommendation. AFS cell administrators MAY use different lists of servers in the AFSDB RRs and DNS SRV RRs if desired for specific effects based on local knowledge of which clients use AFSDB RRs and which clients use DNS SRV RRs. However, AFS servers SHOULD NOT be advertised with AFSDB RRs unless they provide VLDB and

PTS services via UDP on the standard ports. (This falls shy of MUST NOT because it may be useful in some unusual circumstances to advertise via an AFSDB RR a server that provides only one of the two services, but be aware that such a configuration may confuse legacy clients.)
An AFS cell SHOULD have at least one VLDB and at least one PTS server providing service on the standard ports of 7003 and 7002, respectively, since clients without DNS SRV RR support cannot locate servers on non-standard ports.
Clients SHOULD query DNS SRV RRs by default but SHOULD then fall back on AFSDB RRs if no DNS SRV RRs are found. In the absence of DNS SRV RRs, an AFSDB RR of <subtype> 1 SHOULD be treated as equivalent to the following pair of DNS SRV RRs:

```
_afs3-vlserver._udp.<cell> <ttl> IN SRV 0 0 7003 <hostname>
_afs3-prserver._udp.<cell> <ttl> IN SRV 0 0 7002 <hostname>
```

<cell> is the label of the AFSDB RR, <ttl> is its TTL, and <hostname> is the <hostname> value of the AFSDB RR as specified in [RFC1183] (Everhart, C., Mamakos, L., Ullmann, R., and P. Mockapetris, “New DNS RR Definitions,” October 1990.). This is the fully-qualified domain name of the server.

---

6. Example

TOC

The following example includes TCP AFS services, separation of a PTS server from a VLDB server, and use of non-standard ports, all features that either assume future AFS protocol development or are not widely supported by current clients. This example is intended to show the range of possibilities for AFS DNS SRV RRs, not as a practical example for an existing cell. This is a part of the zone file for a fictional example.com domain with AFS services.

```
$ORIGIN example.com.
@                    SOA   dns.example.com. root.example.com. (
                             2009100201 3600 3600 604800 86400 )
                     NS    dns.example.com.
_afs3-vlserver._udp  SRV   0 2 7003 afsdb1.example.com.
_afs3-vlserver._udp  SRV   0 4 7003 afsdb2.example.com.
_afs3-vlserver._udp  SRV   1 0 65500 afsdb3.example.com.

_afs3-vlserver._tcp  SRV   0 0 7003 afsdb3.example.com.

_afs3-prserver._udp  SRV   0 0 7002 afsdb1.example.com.

_afs3-prserver._tcp  SRV   0 0 7002 afsdb3.example.com.

@                    AFSDB 1 afsdb1.example.com.

dns                  A     192.0.2.9
afsdb1               A     192.0.2.10
afsdb2               A     192.0.2.11
afsdb3               A     192.0.2.12
```

In this example, the AFS cell name is example.com.
afsdb1, afsdb2, and afsdb3 all provide VLDB service via UDP. The first two have the same priority but have weights indicating that afsdb1 should get about twice as many clients as afsdb2. afsdb3 should only be used for UDP VLDB service if afsdb1 and afsdb2 are not accessible and provides that service on a non-standard port (65500).
Only one host, afsdb1, provides UDP PTS service.
afsdb3 provides a hypothetical TCP version of AFS VLDB and PTS service on the standard ports and is the only server providing these services over TCP for this cell. Such a TCP-based AFS protocol did not exist at the time this document was written. This example only shows what the record would look like in a hypothetical future if such a protocol were developed.
An AFSDB RR is provided for backward compatibility with older clients. It lists only afsdb1, since only that host provides both VLDB and PTS service over UDP on the standard ports.

---

7. Security Considerations

TOC

Use of DNS SRV RRs for AFS service location poses the same security issues as the existing AFSDB RRs. Specifically, unless the integrity and authenticity of the DNS response are checked, an attacker may forge DNS replies and thereby direct clients at a VLDB or PTS server under the control of the attacker. From there, the attacker may deceive an AFS client about the volumes and file servers in a cell and about the

contents of files and directories in that cell. If the client uses cell data in a trusted way, such as by executing programs out of that AFS cell or using data from the cell as input to other programs, the attacker may be able to further compromise the security of the client and trick it into taking actions under the attacker's control.
This attack can be prevented if the server is authenticated, since the client can then detect a failure to authenticate the attacker's servers and thereby detect possible impersonation. However, this applies only to authenticated AFS access, and much AFS access is unauthenticated. Furthermore, clients after failure to authenticate may fall back to unauthenticated access, which the attacker's servers may permit.
Using an integrity-protected DNS system such as DNSSEC [RFC4033] (Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "DNS Security Introduction and Requirements," March 2005.) can prevent this attack via DNS. However, the underlying vulnerability is inherent in the current AFS protocol and may be exploited in ways other than DNS forgery, such as by forging the results of VLDB queries for an AFS cell. Addressing it properly requires changes to the AFS protocol allowing clients to always authenticate AFS services and discard unauthenticated data. Even in the absence of a DNS system with integrity protection, addition of DNS SRV RRs does not make this vulnerability more severe, only opens another equivalent point of attack.

---

8. IANA Considerations

TOC

This document has no actions for IANA.

---

9. References

TOC

9.1. Normative References

TOC

[RFC1034]	Mockapetris, P., "Domain names - concepts and facilities," STD 13, RFC 1034, November 1987 (TXT).
[RFC1183]	Everhart, C., Mamakos, L., Ullmann, R., and P. Mockapetris, "New DNS RR Definitions," RFC 1183, October 1990 (TXT).
[RFC2119]	Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels," BCP 14, RFC 2119, March 1997 (TXT, HTML, XML).
[RFC2782]	

	Gulbrandsen, A., Vixie, P., and L. Esibov, “A DNS RR for specifying the location of services (DNS SRV),” RFC 2782, February 2000 (TXT).

9.2. Informative References

TOC

[AFS1]	Howard, J., Kazar, M., Menees, S., Nichols, D., Satyanarayanan, M., Sidebotham, R., and M. West, “Scale and Performance in a Distributed File System,” ACM Trans. on Computer Systems 6(1), February 1988.
[AFS2]	Howard, J., “An Overview of the Andrew File System,” CMU-ITC 88-062, February 1988.
[ARLA]	Assar Westerlund, et al., “Arla,” May 2001.
[OPENAFS]	IBM Corporation, et al., “OpenAFS Documentation,” April 2000.
[RFC4033]	Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, “DNS Security Introduction and Requirements,” RFC 4033, March 2005 (TXT).

Author's Address

TOC

	Russ Allbery
	Stanford University
	P.O. Box 20066
	Stanford, CA 94309
	US
Email:	rra@stanford.edu
URI:	http://www.eyrie.org/~eagle/