The ISP Column
A column on various things Internet

February 2024
Geoff Huston

DNS and the DELEG Proposal

The Internet's Domain Name System (DNS) is implemented as a distributed database. The structure of the database mimics the structure of the name space itself, namely a hierarchy where each "node" (or "zone") in the distributed database has a single "parent" node and some number of "child" or descendant nodes (or none), and these linkage points between zones (or “delegations”) are noted in domain names with a period (‘.’) connector between labels (but not exclusively, just to be confusing!).

Parent zones denote a delegation point within the name space (a "zone cut") by using a **NS** Resource Record in the zone data. The value of the NS record is the name of a DNS server that is authoritative for that delegated zone (yes, the specification allows only for a DNS server name, and not its IP address). If a zone is served by multiple authoritative name servers, then multiple NS records are used. If the parent zone is DNSSEC-signed, then these delegation NS records are unsigned. Prior to DNSSEC, the only resource record in the parent zone for this delegation point was the NS record. All other data about this name was found in the child zone.

If the parent zone is DNSSEC-signed, then there will also be an NSEC (or NSEC3) record and the NSEC record's associated RRSIG signature, and if the delegated zone is DNSSEC-signed, then the parent zone will also hold a DS record and its associated RRSIG record.

The names of these name servers are not enough to allow a resolver to query these servers, as we also need to know their IP addresses. If the name servers are named by a name that exists within the child zone, then there is a circular dependency that cannot be resolved. To aid resolvers and speed up the name resolution process, parent zones also usually carry "Glue Records," which are the IP addresses of the nameservers. Such glue records are used in the Additional Section of referral responses, providing the querying resolver with the wherewithal to query the delegated zone.

These Glue Records have been the source of some issues with DNS resolution and have been misused in the past to try and corrupt the cache DNS information that is held by a resolver. They are not “linked” to zone from which the names have been drawn and may be incorrect. They are also not DNSSEC-signed.

When the name server name is not drawn from the delegated zone (and its descendants), or in DNS-terminology is not “in-balliwick” the glue record is not strictly required, as the resolver can generally resolve the name using the DNS resolution mechanisms recursively (name loops not withstanding).

However, resolving name server records slows down name resolution, and as long as these names are only used in the context of the top-domain zone traversal, glue records are generally regarded as a useful part of DNS name resolution.

An NS record that is located at the apex of a child zone does not denote a further delegation but is intended to also list the authoritative name servers for that zone. This list of authoritative name servers in the child zone should match the list that is held in the parent zone. If the zone is DNSSEC-signed, then these NS records at the apex of the child zone will be signed by the child zone's zone-signing key, as the child zone that is authoritative for the names of the name servers that serve this zone.

In summary, the links within the DNS that maintain the cohesion of this distributed data framework are these NS records, which are duplicated across both sides of each delegation point.

> "As the last installation step, the delegation NS RRs and glue RRs necessary to make the delegation effective should be added to the parent zone. The administrators of both zones should insure that the NS and glue RRs which mark both sides of the cut are consistent and remain so."
> RFC 1034, sec 4.2.2

The child zone is properly the "correct" party to determine which name servers serve the content of this zone, and the child zone is authoritative fort this information. However, in the DNS top-down name resolution process the parent zone must inform the querier that it is not authoritative for the name because a zone cut exists and pass back a referral to the nameservers for the zone. It's therefore the parent's copy of the nameserver information that is used by resolvers when they receive a referral response.

Here is an example of a referral response from a parent zone nameserver:

```
$ dig +norecurse A www.potaroo.net @a.gtld-servers.net.

; <<>> DiG 9.18.21 <<>> +norecurse www.potaroo.net @a.gtld-servers.net.
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 58995
;; flags: qr; QUERY: 1, ANSWER: 0, AUTHORITY: 2, ADDITIONAL: 5

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 4096
;; QUESTION SECTION:
;www.potaroo.net.              IN      A

;; AUTHORITY SECTION:
potaroo.net.           172800  IN      NS      ns2.potaroo.net.
potaroo.net.           172800  IN      NS      ns1.potaroo.net.

;; ADDITIONAL SECTION:
ns2.potaroo.net.       172800  IN      A       203.133.248.6
ns2.potaroo.net.       172800  IN      AAAA    2401:2000:6660::6
ns1.potaroo.net.       172800  IN      A       203.133.248.2
ns1.potaroo.net.       172800  IN      AAAA    2401:2000:6660::2
```

In the context of resolving the DNS name `www.potaroo.net`, the resolver will now be able to pass the same DNS query to either of these name servers for the zone `potaroo.net`, and this referral response has provided the IP addresses of the name servers for this query.

This structure has served the DNS adequately for many decades, so why change it?

Changing the DNS Mechanisms for Delegation

There are a number of pressures for change here. One source of such pressure is that this referral response is not protected with DNSSEC signatures, and is accordingly prone to on-path substitution attacks.

As Joe Abley explains in a draft, dnsop-refer:

> "A Standard Referral response from an authoritative DNS server includes an NS RRset. It is not possible for the response to include a corresponding RRSIG RRset, since the administrator of a parent zone is generally not in possession of the private keys needed to make signatures in a child zone. The lack of signatures means that the Standard Referral response is subject to on-path substitution attacks,

> even if both parent and child zones are signed and the originator of the request that triggered the referral response requests DNSSEC data (with DO=1) and is capable of validating responses."
> https://datatracker.ietf.org/doc/html/draft-jabley-dnsop-refer-00

Of course, there is a simpler way to confirm the authenticity of the NS records provided in a referral response, if the delegated zone is signed, and that is to use these name servers and ask the child zone for its signed NS record(s), as proposed in a draft on delegation revalidation. (https://datatracker.ietf.org/doc/html/draft-ietf-dnsop-ns-revalidation-00)

A different approach was proposed by Kazunori Fujiwara in the draft delegation-information-signer (https://datatracker.ietf.org/doc/html/draft-fujiwara-dnsop-delegation-information-signer-00), where the delegation information is hashed and signed within the parent zone, allowing a querier to retrieve and validate this hash value and confirm that it is the hash of the delegation information.

However, this is not the only pressure on the delegation mechanism. A second pressure is the incomplete work in the DNS Privacy area. The use of encrypted transports, namely DNS over TLS, HTTPS and QUIC have been defined by the IETF's DPRIVE Working Group, but the mechanisms to trigger the use of these encrypted protocols is unclear. For stub resolvers querying recursive resolvers, the anticipated future query load is large enough that the overhead of probing the recursive resolver for its transport capabilities can be amortised over subsequent queries, but the same consideration does not necessarily apply to querying authoritative resolvers.

The use of Service Binding records in the DNS offers a mechanism for a client to efficiently discover the capabilities of a server. The IETF published RFC 9460, Service Binding and Parameter Specification via the DNS (SVCB and HTTPS Resource Records) in November 2023, which defines a DNS resource record that is associated with a service name that not only provides the IP address(es) of the service point, but also can define the transport protocol, port value, and aliases. This approach is sufficiently generic that it can be used for many service types, including the DNS itself, and at the same time the IETF published RFC9461, Service Binding Mapping for DNS servers:

> "The SVCB DNS resource record type expresses a bound collection of endpoint metadata, for use when establishing a connection to a named service. DNS itself can be such a service, when the server is identified by a domain name. This document provides the SVCB mapping for named DNS servers, allowing them to indicate support for encrypted transport protocols."
> RFC 9461 - https://datatracker.ietf.org/doc/html/rfc9461

Using the SVCB service record its possible for an authoritative DNS server of a zone to indicate that queries should be made over a non-default port, or can nominate the transport protocol via the "ALPN" key value. In the following example, the server `ns1.potaroo.net` could indicate its support for queries using DNS over TLS with the record:

```
_dns.ns1.potaroo.net. 172800 IN SVCB 1 ns1.potaroo.net. (
    alpn=dot ipv4hint=203.133.248.2 ipv6hint=2401:2000:6660::2)
```

As RFC9461 also points out "nothing in this document indicates whether the service is intended for use as a recursive or authoritative DNS server."

The record exists in the context of the zone in which the name is defined, so it can be signed and potentially secure. In our example, the label `_dns.ns1.potaroo.net.` is defined in the zone `potaroo.net`, and as potaroo.net is DNSSEC-signed, then the record is signed by the zone signing key for this zone. The parent zone is not necessarily aware of this record, and still provides the simple NS records and associated glue. A resolver would still encounter these NS records on first use, even when attempting to resolve a query for the SVCB record for `_dns.ns1.potaroo.net` it would still have to use DNS over UDP port 53 to make this query for the first time, as implicitly defined by the parent-served NS record. The resolver may then choose to query for a DNS Service record for this name server. The question is whether the semantics of the DNS Service Binding record could be merged with the name delegation function of the NS record.

So, how could we combine the DNS service SVCB profile with name server functionality to the extent the combined data of name server and service profile could be used in referral responses?

The DELEG Resource Record

One approach is to use a new RR at the parent which can be used in referral responses that combines the semantics of delegation and the listing the name server name and the IP addresses of this service in the current NS and Glue records, but also adding the transport protocol capabilities, and combine all of this information in a single RR. This is the intention of the proposed **DELEG** Resource Record (https://datatracker.ietf.org/doc/draft-dnsop-deleg/). This proposal borrows the semantics of the RFC 9461 DNS SVCB record, but uses the delegation label rather than the service name and the binding point in the DNS, and builds upon an earlier proposal for NS2 and NST records, that were described in an earlier draft by Tim April (https://datatracker.ietf.org/doc/html/draft-tapril-ns2-01).

Taking our example delegated domain `potaroo.net`, and the DNS service binding example from above, a comparable DELEG record would be:

```
  potaroo.net. 172800 IN DELEG 1 ns1.potaroo.net. (
       alpn=dot ipv4hint=203.133.248.2 ipv6hint=2401:2000:6660::2)
```

This record indicates that `potaroo.net` is a delegation point in the `net` zone, and the server should be queried using DNS over TLS on the default TCP port 853. As the `net` zone is a DNSSEC-signed zone, this record would also have an associated RRSIG signature record.

When used as a referral response, the parent server is not necessarily aware if the querier is aware of the DELEG record, so simply responding with this response as a referral response would be ineffectual if the querier was not aware of the use of the DELEG record.

One possible solution to this would be for the querier to include an EDNS(0) option in the query to indicate that the querier is capable of processing a DELEG referral response (in a similar manner to the description of querier capabilities of the earlier work on the REFER response).

Another approach is to fold in backward compatibility with the existing referral mechanism. The server would assemble a referral response that includes both the DELEG records and the NS records in the Authority Section, and the unsigned glue records in the Additional Section. If the EDNS(0) DO bit was set in the query, then the RRSIG for the DELEG record would also be included in the response. The draft specification of the DELEG record also specifies that the referral response would also include the DS RRset and its associated RRSIG signature. A referral response for our example DNSSEC-signed delegation would be as follows:

```
  potaroo.net 172800 IN DELEG 1 ns1.potaroo.net. (
      alpn=dot ipv4hint=203.133.248.2 ipv6hint=2401:2000:6660::2)
  potaroo.net 172800 IN DELEG 2 ns1.potaroo.net. (
      ipv4hint=203.133.248.2 ipv6hint=2401:2000:6660::2)
  potaroo.net 172800 IN DELEG 2 ns2.potaroo.net. (
      ipv4hint=203.133.248.6 ipv6hint=2401:2000:6660::6)
  potaroo.net.        37917   IN     RRSIG   DELEG 13 2 86400 20240208082017 ...

  potaroo.net.        37917   IN     DS      43552 13 1 BA715ACF7E2F501F05FB4E89B22A6F032327E430
  potaroo.net.        37917   IN     RRSIG   DS 13 2 86400 20240208082017 ...

  potaroo.net.        172800  IN     NS      ns2.potaroo.net.
  potaroo.net.        172800  IN     NS      ns1.potaroo.net.


  ns2.potaroo.net.    172800  IN     A       203.133.248.6
  ns2.potaroo.net.    172800  IN     AAAA    2401:2000:6660::6
  ns1.potaroo.net.    172800  IN     A       203.133.248.2
  ns1.potaroo.net.    172800  IN     AAAA    2401:2000:6660::2
```

If the ability to nominate additional DNS transport protocols is all the DELEG RR is good for, then I think it's highly unlikely that it will gain much acceptance in the DNS environment. Any change to the domain name provisioning imposes additional cost, and its challenging to quantify the benefits that would balance against the cost of implementing the necessary change for this particular resource record. In theory, a parent zone operator could synthesise a DELEG record from the existing NS and associated glue information, but as the response would already contain the NS and glue records, then it's reasonable to ask why this additional record in the referral response would be of any value if it's a simple syntactic repackaging of the referral information. The community of users who place tangible value on using an encrypted channel between the recursive resolver and the authoritative server is not exactly large. (Don't forget that the IP identity of the original querier is not present is such queries.)

A signed DELEG record would allow a querier to validate the referral and detect potential attacks that attempt to perform substitution in the referral response. This is pushing DNSSEC into a new space. At present DNSSEC will provide a validating resolver some assurance that the response they received to their query is authentic and current. This assurance relates only to the final response itself, and not the "correctness" of the sequence of name servers used to discover the authoritative nameservers for the zone. Signing these DELEG records would allow a validator to provide some assurance as to the authenticity of these referral responses, but this would only be at the cost of additional validation queries, particularly for the DNSKEY records of the zone in which the DELEG record is located, and other queries if the DELEG record is an alias record that leads to a different validation path. A validating client being led astray would presumably detect the attempted attack when the response fails validation in any case, so the time saved by performing validation checks on the DELEG records and stopping as soon as the DELEG record fails validation would be offset of the additional time spent on assembling the validation information for each of these delegations.

However, there is one aspect of the DELEG record that appears to offer some benefit over the existing NS referral functionality. The service hosting industry relies heavily on the DNS CNAME structure. CNAMEs allow individual names to be "lifted" out of the original locus of administrative control and shifted over to the control of the service hosting entity. For example, the domain name `www.apple.com` is provisioned using Akamai, and the DNS name `www.apple.com` is passed across to Akamai's control using the CNAME alias transform:

```
$ dig A www.apple.com
www.apple.com.    1750      IN      CNAME   www.apple.com.edgekey.net.
www.apple.com.edgekey.net. 13319 IN CNAME   www.apple.com.edgekey.net.globalredir.akadns.net.
www.apple.com.edgekey.net.globalredir.akadns.net. 2069 IN CNAME   e6858.dscx.akamaiedge.net.
e6858.dscx.akamaiedge.net. 10       IN      A       23.204.64.212
```

If CNAME RRs are so useful for content service hosting, then why not use CNAMEs for NS records to facilitate DNS service hosting? The current DNS specification stipulates that CNAMES cannot be used for Nameserver records:

> "The domain name used as the value of a NS resource record, or part of the value of a MX resource record must not be an alias. Not only is the specification clear on this point, but using an alias in either of these positions neither works as well as might be hoped, nor well fulfills the ambition that may have led to this approach. This domain name must have as its value one or more address records. Currently those will be A records, however in the future other record types giving addressing information may be acceptable. It can also have other RRs, but never a CNAME RR."
> RFC 2181, section 10.3

The SVCB specification allows for Alias Mode (Sec 2.4.2, RFC94360). It also allows for Alias Mode SVCB records to exist at the zone apex, where CNAME records are not permitted. The DELEG specification allows for the equivalent of Alias Mode name server records. For example:

The `.net` zone is contains alias model DELEG record:

```
potaroo.net 172800 IN DELEG 0 potaroo.net.dnshosting.example.com
```

The `dnshosting.example.com` zone can contain the SVCB record for the dns hosting service:

```
potaroo.net.dnshosting.example.com. 3600 IN SVCB . (
    alpn=dot ipv4hint=203.133.248.2 ipv6hint=2401:2000:6660::2)
```

This has a limitation in that the alias mode DELEG record contains no other attributes, so it is not possible to alias into an "in-domain" name (a name in the same zone as the zone that is being delegated, or a descendent zone to avoid circular dependencies - RFC8499, Sec. 7).

Like a CNAME, this construct can be used to shift the administrative control of a name server to a DNS operator, who can then make changes to the name server without any further need to update the parent zone. How to handle the backward compatible glue records for an alias mode record is an open issue. Presumably if the alias refers to a zone that is unrelated to the parent zone, then the backward compatible NS records can refer to the alias name and the glue records can be omitted, leaving it to the resolver client to resolve the SVCB alias name.

Whether this additional alias behaviour in delegation is enough to impel the DELEG record into a mode of general adoption remains to be seen. The observation is that the DNS name industry has worked for some decades within the current framework of delegation, and the changes being contemplated with the DELEG record do not seem to offer any fundamental new efficiencies or capabilities that would motivate its general adoption.

What this proposed DELEG record does not do is alter the inter-zone communication requirement. As the draft specification notes:

> “Construction of a DELEG RR requires knowledge which implies communication between the operators of the child and parent zones. This communication is an operational matter not covered by this document.”
> https://datatracker.ietf.org/doc/draft-dnsop-deleg/

Where the Extensible Provisioning Protocol (EPP) (RFC 3731) is used to pass configuration parameters between registrars and registries, then the EPP process would need to be extended to allow the specification of a DELEG record to be included.

An alternative for DNSSEC-signed zones is to use the Child-to-Parent Synchronisation mechanism (RFC 7477) and add the DELEG record to the child’s signed CSYNC record for the parent to pick up this record, validate it, and then add it to the parent zone with the parent’s signature. How the parent zone operator is to detect a change in this record is a topic in its own right!

Observations

All large engineered systems accrete stasis and resist change over time. This implies that changes to very large systems become more like piecemeal local customisations that need to coexist with the general status quo.

The DELEG record is still in its early stages of consideration and it’s not even clear whether this approach will garner widespread support or not.

DELEG attempts to address three shortcomings of the current DNS delegation process.

The first is that the parent’s presentation of the NS records to a resolver is unsigned and is therefore prone to various forms of substitution attacks. DELEG, like a number of other proposals, attempts to address this by proposing a new delegation resource record that is authoritative in the parent zone, so that can be served with a DNSSEC signature. This is an extension to DNSSEC-signing in the DNS, where it’s not just the DNS response that it being authenticated, but the resolution path being used. The marginal benefit of this additional function is difficult to assess, given that a resolution path that misleads the recursive resolver will more than likely end with a response that will fail DNSSEC validation in any case.

The second is an inability to specify an alternative transport protocol to use to query the authoritative servers of the delegated zone. Again, the general benefit of this is unclear. The encrypted channel transport protocols need to perform an initial packet exchange to set up the shared encryption state. In the stub-to-recursive resolver situation this overhead can be offset by the subsequent queries that use the established channel, and there is a marginal gain in avoiding the delays in switching over from UDP to TCP for large responses. However, this is not necessarily the case in the recursive-to-authoritative scenario, as a single recursive resolver may query a large number of authoritative servers in a relatively short span of time.

It's the final shortcoming where the DELEG resource record may offer a useful function. The inability to use a CNAME record as a target of an NS record means that it's hard to move a DNS name server name away from the control of the child zone operator. If a DNS operator wishes to dynamically generate name server records that can perform load balancing or offer records that are located close to the querier to improve DNS performance, then this is practically impossible in the current NS record structure.

This looks promising, but there is the residual consideration of the backward compatibility of the proposed DELEG structure. Do the NS and glue records that are also packaged in the delegation response have to match the DELEG record contents? Or do the DELEG records specify additional name servers that may be queried in a higher level of preference to the servers listed in the NS records? Do we need this somewhat clumsy stuffing of referral responses with both DELEG and NS records in any case? Why not just borrow from the earlier REFER proposal and respond with only DELEG records if the query indicates via an EDNS(0) option setting that it can handle DELEG referrals? Otherwise the server would respond with NS and Glue records as it does today to indicate a zone cut and referral.

The only material about the DELEG records is posted in a first version of an Internet draft, and it's highly likely that the material will be further refined in the coming months. Or maybe even dropped completely!

T. April, P. Špaček, R. Weber, and D. Lawrence, "Extensible Delegation for DNS", Work in progress: draft-dnsop-deleg-00, January 2024. https://datatracker.ietf.org/doc/draft-dnsop-deleg/

Disclaimer

The above views do not necessarily represent the views or positions of the Asia Pacific Network Information Centre.

Author

Geoff Huston AM, B.Sc., M.Sc., is the Chief Scientist at APNIC, the Regional Internet Registry serving the Asia Pacific region.

www.potaroo.net