dnsext	C. Contavalli		TOC
Internet-Draft	W. van der Gaast	
Intended status: Experimental	Google	
Expires: November 22, 2010	S. Leach	
	Name.com	
	D. Rodden	
	Neustar	
	May 21, 2010	

Client IP information in DNS requests
draft-vandergaast-edns-client-ip-01

Abstract

This draft defines an EDNS0 extension to carry relevant network range information. In a query, it conveys the network address of the originator. In a response, it conveys the scope of network addresses that the answer is intended.

Status of this Memo

Copyright Notice

Table of Contents

---

1.   Introduction

Many Authoritative nameservers today return different replies based on
the perceived topological location of the user. These servers use the
IP address of the incoming query to identify that location. Since most
queries come from intermediate recursive resolvers, the source address
is that of the recursive rather than of the query originator.
Traditionally and probably still in the majority of instances,
recursive resolvers are reasonably close in the topological sense to
the stub resolvers or forwarders that are the source of queries. For
these resolvers, using their own IP address is sufficient for authority
servers that tailor responses based upon location of the querier.
Increasingly though a class of remote recursive servers has arisen that
serves query sources without regard to topology. The motivation for a
query source to use a remote recursive server varies but is usually

because of some enhanced experience, such as greater cache security or applying policies regarding where users may connect. (Although political censorship usually comes to mind here, the same actions may be used by a parent when setting controls on where a minor may connect.) When using a remote recursive server, there can no longer be any assumption of close proximity between the originator and the recursive, leading to less than optimal replies from the authority servers.

A similar situation exists within some ISPs where the recursive servers are topologically distant from some edges of the ISP network, resulting in less than optimal replies from the authority servers.

This draft defines an EDNS0 option to convey network information that is relevant to the message but not otherwise included in the datagram. This will provide the mechanism to carry sufficient network information about the originator for the authority server to tailor responses. It also provides for the authority server to indicate the scope of network addresses that the tailored answer is intended. This EDNS0 option is intended for those recursive and authority servers that would benefit from the extension and not for general purpose deployment. It is completely optional and can safely be ignored by servers that choose not to implement it or enable it.

This draft also includes guidelines on how to best cache those results and provides recommendations on when this protocol extension should be used.

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

---

2. Terminology

Stub Resolver: A simple DNS protocol implementation on the client side as described in [RFC1034] (Mockapetris, P., "Domain names - concepts and facilities," November 1987.) section 5.3.1.

Authoritative Nameserver: A nameserver that has authority over one or more DNS zones. These are normally not contacted by clients directly but by Recursive Resolvers. Described in [RFC1035] (Mockapetris, P., "Domain names - implementation and specification," November 1987.) chapter 6.

Recursive Resolver:

A nameserver that is responsible for resolving domain names for clients by following the domain's delegation chain, starting at the root. Recursive Resolvers frequently use caches to be able to respond to client queries quickly. Described in [RFC1035] (Mockapetris, P., "Domain names - implementation and specification," November 1987.) chapter 7.

Intermediate Nameserver: Any nameserver (possibly a Recursive Resolver) in between the Stub Resolver and the Authoritative Nameserver.

Third-party Nameserver: Recursive Resolvers provided by parties that are not Internet Service Providers (ISPs). These services are often offered as substitutes for ISP-run nameservers.

Optimized reply: A reply from a nameserver that is optimized for the node that sent the request, normally based on performance (i.e. lowest latency, least number of hops, topological distance, ...).

Topologically close: Refers to two hosts being close in terms of number of hops or time it takes for a packet to travel from one host to the other. The concept of topological distance is only loosely related to the concept of geographical distance: two geographically close hosts can still be very distant from a topological perspective.

---

3. Option format

This draft uses an EDNS0 ([RFC2671] (Vixie, P., "Extension Mechanisms for DNS (EDNS0)," August 1999.)) option to include client IP information in DNS messages. The option is structured as follows:

```
                +0 (MSB)                          +1 (LSB)
        +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
    0:  |                        OPTION-CODE                           |
        +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
    2:  |                       OPTION-LENGTH                          |
        +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
    4:  |                          FAMILY                              |
        +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
    6:  |       SOURCE NETMASK       |        SCOPE NETMASK            |
        +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
    7:  |                         ADDRESS...                           /
        +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
```

  *(Defined in [RFC2671] (Vixie, P., "Extension Mechanisms for DNS
   (EDNS0)," August 1999.)) OPTION-CODE, 2 octets, for edns-client-
   subnet is TBD.

  *(Defined in [RFC2671] (Vixie, P., "Extension Mechanisms for DNS
   (EDNS0)," August 1999.)) OPTION-LENGTH, 2 octets, contains the
   length of the payload (everything after OPTION-LENGTH) in bytes.

  *FAMILY, 2 octets, indicates the family of the address contained
   in the option, using address family codes as assigned by IANA in
   IANA-AFI.

The format of the address part depends on the value of FAMILY. This
document only defines the format for FAMILY 1 (IP version 4) and 2 (IP
version 6), which are as follows:

  *SOURCE NETMASK, 1 octet, in requests, indicates how many most-
   significant bits of the SOURCE ADDRESS are included (i.e. a
   netmask in CIDR notation). In replies, it echoes back the value
   from the query.

  *SCOPE NETMASK, 1 octet, in requests, it MUST be set to 0. In
   replies, it indicates for which supernets of ADDRESS the reply
   can be cached, or which netmask would be necessary in requests to
   make a better choice, see next few sections.

  *ADDRESS, variable number of octets, contains either an IPv4 or
   IPv6 address (depending on FAMILY), truncated to the number of
   bits indicated by the SOURCE NETMASK field, with bits set to 0 to
   pad up to the end of the last octet used.

All fields are in network byte order.

4. Protocol description

The edns-client-subnet extension allows DNS servers to propagate the network address of the client that initiated the resolution through to Authoritative Nameservers during recursion.
Servers that receive queries containing an edns-client-subnet option can generate answers based on the original network address of the client. Those answers will generally be optimized for that client and other clients in the same network.
The option also allows Authoritative Nameservers to specify the network range for which the reply can be cached and re-used.

4.1. Originating the option

The edns-client-subnet option can be added by Intermediate Nameservers or Stub Resolvers. If an Intermediate Nameserver receives a query from a routable (i.e. not private IP space as described in [RFC1918] (Rekhter, Y., Moskowitz, R., Karrenberg, D., Groot, G., and E. Lear, "Address Allocation for Private Internets," February 1996.)) IP address and it doesn't yet have the option, it MAY add an edns-client-subnet option populated with the source IP address of the query.
Alternatively, a Stub Resolver MAY generate DNS queries with an edns-client-subnet option, for example if it has better knowledge of where the connection following the DNS lookup is going to enter the public network, or to request anonymization by including an edns-client-subnet option with the address 0.0.0.0/0.
If an Intermediate Nameserver supporting edns-client-subnet receives a query that already has a valid edns-client-subnet option, this option MUST be passed through as-is and MUST NOT be modified.
For privacy reasons, and because the whole IP address is rarely required to determine an optimized reply, the ADDRESS field in the option SHOULD be truncated to a certain number of bits, chosen by the administrators of the server, as described in Section 8 (Security Considerations).
Intermediate Nameservers that have not implemented or enabled support for the edns-client-subnet can safely ignore the option within incoming queries. Such a server MUST NOT include an edns-client-subnet option within replies to indicate lack of support for the option.
A Recursive Resolver MAY be configured to not include (or drop an existing) edns-client-subnet option completely when querying Authoritative Nameservers from which a delegation response is expected, for example TLD servers or root servers.

4.2. Generating a response

When a query containing an edns-client-subnet option is received, an Authoritative Nameserver supporting edns-client-subnet MAY use the address information specified in the option in order to generate an optimized reply.
Authoritative servers that have not implemented or enabled support for the edns-client-subnet may safely ignore the option within incoming queries. Such a server MUST NOT include an edns-client-subnet option within replies to indicate lack of support for the option.
Requests with an edns-client-subnet option considered invalid (i.e. wrong formatting, unsupported address family, private address space) MUST be treated as if no edns-client-subnet option was specified.
If the Authoritative Nameserver decides to use information from the edns-client-subnet option to calculate a response, it MUST include the option in the response to indicate that the information was used (and has to be cached accordingly). If the option was not included in a query, it MUST NOT be included in the response.
The FAMILY, ADDRESS and SOURCE NETMASK in the response MUST match those in the request. Echoing back the address and netmask helps to mitigate certain attack vectors, as described in Section 8 (Security Considerations).
The SCOPE NETMASK in the reply indicates the network range that the answer is intended for.
A SCOPE NETMASK value larger than the SOURCE NETMASK indicates that the address range provided in the query was not specific enough to select a single, best response, and that an optimal reply would require at least SCOPE NETMASK bits of address information.
Conversely, a lower SCOPE NETMASK indicates that more bits than necessary were provided.
In both cases, the value of the SCOPE NETMASK in the reply has strong implications with regard to how the reply will be cached by Intermediate Nameservers, as described in Section 4.3 (Handling edns-client-subnet replies and caching).
If the edns-client-subnet option in the request is not used at all (for example if an optimized reply was temporarily unavailable or not supported for the requested domain name), a server supporting edns-client-subnet MUST indicate that no bits of the ADDRESS in the request have been used by specifying a SCOPE NETMASK of 0 (equivalent to the networks 0.0.0.0/0 or ::/0).
If no optimized answer could be found at all for the ADDRESS and SOURCE NETMASK indicated in the query, the Authoritative Nameserver SHOULD still return the best result it knows of (i.e. by using the query source IP address instead, or a sensible default), and indicate that this result should only be cached for the FAMILY, ADDRESS and SOURCE NETMASK indicated in the request. The server will indicate this by copying the SOURCE NETMASK into the SCOPE NETMASK field.

4.3. Handling edns-client-subnet replies and caching

When an Intermediate Nameserver receives a reply containing an edns-client-subnet option, it will return a reply to its client and may cache the result.
If the FAMILY, ADDRESS and SOURCE NETMASK fields in the reply don't match the fields in the corresponding request, the full reply MUST be dropped, as described in Section 8 (Security Considerations).
In the cache, any resource record in the answer section will be tied to the network specified by the FAMILY, ADDRESS and SCOPE NETMASK fields, as detailed below.
If another query is received matching the entry in the cache, the resolver will verify that the FAMILY and ADDRESS that represent the client match any of the networks in the cache for that entry.
If the address of the client is within any of the networks in the cache, then the cached response MUST be returned as usual. In case the address of the client matches multiple networks in the cache, the entry with the highest SCOPE NETMASK value MUST be returned, as with most route-matching algorithms.
If the address of the client does not match any network in the cache, then the Recursive Resolver MUST behave as if no match was found and perform resolution as usual. This is necessary to avoid sub-optimal replies in the cache from being returned to the wrong clients, and to avoid a single request coming from a client on a different network from polluting the cache with a sub-optimal reply for all the users of that resolver.
Note that every time a Recursive Resolver queries an Authoritative Nameserver by forwarding the edns-client-subnet option that it received from another client, a low SOURCE NETMASK in the original request could cause a sub-optimal reply to be returned by the Authoritative Nameserver.
To avoid this sub-optimal reply from being served from cache for clients for which a better reply would be available, the Recursive Resolver MUST check the SCOPE NETMASK that was returned by the Authoritative Nameserver:

   *If the SCOPE NETMASK in the reply is higher than the SOURCE NETMASK, it means that the reply might be sub-optimal. A Recursive Resolver MUST return this entry from cache only to queries that do not contain or allow a higher SOURCE NETMASK to be forwarded.

   *If the SCOPE NETMASK in the reply is lower or equal to the SOURCE NETMASK, the reply is optimal, and SHOULD be returned from cache to any client within the network range indicated by ADDRESS and SCOPE NETMASK.

When another request is performed, the existing entries SHOULD be kept in the cache until their TTL expires, as per standard behavior.
As another reply is received, the reply will be tied to a different network. The server MAY keep in cache both replies, and return the most appropriate one depending on the address of the client.
Any reply containing an edns-client-subnet option considered invalid should be treated as if no edns-client-subnet option was specified at all.
Replies coming from servers not supporting edns-client-subnet or otherwise not containing an edns-client-subnet option SHOULD be considered as containing a SCOPE NETMASK of 0 (e.g., cache the result for 0.0.0.0/0 or ::/0) for all the supported families.
In any case, the response from the resolver to the client MUST NOT contain the edns-client-subnet option if none was present in the client's original request. If the original client request contained a valid edns-client-subnet option that was used during recursion, the Recursive Resolver MUST include the edns-client-subnet option from the Authoritative Nameserver response in the response to the client.
Enabling support for edns-client-subnet in a recursive resolver will significantly increase the size of the cache, reduce the number of results that can be served from cache, and increase the load on the server. Implementing the mitigation techniques described in Section 8 (Security Considerations) is strongly recommended.

---

5. IANA Considerations

We request IANA to assign an option code for edns-client-subnet, as specified in [RFC2671] (Vixie, P., "Extension Mechanisms for DNS (EDNS0)," August 1999.). Within this document, the text 'TBD' should be replaced with the option code assigned by IANA.

---

6. DNSSEC Considerations

The presence or absence of an OPT resource record containing an edns-client-subnet option in a DNS query does not change the usage of those resource records and mechanisms used to provide data origin authentication and data integrity to the DNS, as described in [RFC4033] (Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "DNS Security Introduction and Requirements," March 2005.), [RFC4034] (Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "Resource Records for the DNS Security Extensions," March 2005.) and [RFC4035] (Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "Protocol Modifications for the DNS Security Extensions," March 2005.).

7. NAT Considerations

Special awareness of edns-client-subnet in devices that perform NAT as described in [RFC2663] (Srisuresh, P. and M. Holdrege, "IP Network Address Translator (NAT) Terminology and Considerations," .) is not required, queries can be passed through as-is. The client's network address MUST NOT be added, and existing edns-client-subnet options, if present, MUST NOT be modified by NAT devices.
Recursive Resolvers sited behind NAT devices MUST NOT add their external network address in an edns-client-subnet options, and MUST behave exactly as described in the previous sections.
Note that Authoritative Nameservers or Recursive Resolvers can still provide an optimized reply by looking at the source IP of the query.

8. Security Considerations

8.1. Privacy

With the edns-client-subnet option, the network address of the client that initiated the resolution becomes visible to all servers involved in the resolution process. Additionally, it will be visible from any network traversed by the DNS packets.
To protect users' privacy, Recursive Resolvers are strongly encouraged to conceal part of the IP address of the user by truncating IPv4 addresses to 24 bits. No recommendation is provided for IPv6 at this time, but IPv6 addresses should be similarly truncated in order to not allow to uniquely identify the client.
Users who wish their full IP address to be hidden can include an edns-client-subnet option specifying the wildcard address 0.0.0.0/0 (i.e. FAMILY set to 1 (IPv4), SOURCE NETMASK to 0 and no ADDRESS). As described in previous sections, this option will be forwarded across all the Recursive Resolvers supporting edns-client-subnet, which MUST NOT modify it to include the network address of the client.
Note that even without edns-client-subnet options, any server queried directly by the user will be able to see the full client IP address.
Recursive Resolvers or Authoritative Nameservers MAY use the source IP address of requests to return a cached entry or to generate an optimized reply that best matches the request.

8.2. Birthday attacks

edns-client-subnet adds information to the q-tuple. This allows an
attacker to send a caching Intermediate Nameserver multiple queries
with spoofed IP addresses either in the edns-client-subnet option or as
the source IP. These queries will trigger multiple outgoing queries
with the same name, type and class, just different address information
in the edns-client-subnet option.
With multiple queries for the same name in flight, the attacker has a
higher chance of success in sending a matching response (with the
address 0.0.0.0/0 to still get it cached for many hosts).
To counter this, every edns-client-subnet option in a response packet
MUST contain the full FAMILY, ADDRESS and SOURCE NETMASK fields from
the corresponding request. Intermediate Nameservers processing a
response MUST verify that these match, and MUST discard the entire
reply if they do not.

8.3. Cache pollution

It is simple for an arbitrary resolver or client to provide false
information in the edns-client-subnet option, or to send UDP packets
with forged source IP addresses.
This could be used to pollute the cache of intermediate resolvers, by
filling it with results that will rarely (if ever) be used, or to
reverse engineer the algorithms (or data) used by the Authoritative
Nameserver to calculate the optimized answers.
Even without malicious intent, third-party Recursive Resolvers
providing answers to clients in multiple networks will need to cache
different replies for different networks, putting more pressure on the
cache.
To mitigate those problems:

*Recursive Resolvers implementing edns-client-subnet should only
 enable it in deployments where it is expected to bring clear
 advantages to the end users. For example, when expecting clients
 from a variety of networks or from a wide geographical area. Due
 to the high cache pressure introduced by edns-client-subnet, the
 feature must be disabled in all default configurations.

*Recursive Resolvers should limit the number of networks and
 answers they keep in the cache for a given query.

*Recursive Resolvers should limit the number of total different
 networks that they keep in cache.

*Recursive Resolvers should never send edns-client-subnet options with SOURCE NETMASKs providing more bits in the ADDRESS than they are willing to cache responses for.

*Recursive Resolvers should implement algorithms to improve the cache hit rate, given the size constraints indicated above. Recursive Resolvers may, for example, decide to discard more specific cache entries first.

*Authoritative Nameservers and Recursive Resolvers should discard known to be wrong or known to be forged edns-client-subnet options. They must at least ignore unroutable addresses, including the ones defined in [RFC1918] (Rekhter, Y., Moskowitz, R., Karrenberg, D., Groot, G., and E. Lear, "Address Allocation for Private Internets," February 1996.) and [RFC4193] (Hinden, R. and B. Haberman, "Unique Local IPv6 Unicast Addresses," October 2005.), and should ignore and never forward edns-client-subnet options specifying networks or addresses that are known not to be served by those servers when feasible.

*Authoritative Nameservers consider the edns-client-subnet option just as a hint to provide better results. They can decide to ignore the content of the edns-client-subnet option based on black or white lists, rate limiting mechanisms, or any other logic implemented in the software.

---

9. Example

1. A stub resolver SR with IP address 192.0.2.37 tries to resolve www.example.com, by forwarding the query to the Recursive Resolver R from IP address IP, asking for recursion.

2. R, supporting edns-client-subnet, looks up www.example.com in its cache. An entry is found neither for www.example.com, nor for example.com.

3. R builds a query to send to the root and .com servers. The implementation of R does not include an edns-client-subnet option when querying TLD or root nameservers, because there is no expectation to receive a client-network-specific response. Thus, no edns-client-subnet option is added, and resolution is performed as usual.

4. R now knows the Authoritative Nameserver ANS responsible for example.com.

5. R prepares a new query for www.example.com, including an edns-client-subnet option with:

    *OPTION-CODE, set to TBD.

    *OPTION-LENGTH, set to 0x00 0x07.

    *FAMILY, set to 0x00 0x01 as IP is an IPv4 address.

    *SOURCE NETMASK, set to 0x18, as R is configured to conceal the last 8 bits of every IPv4 address.

    *SCOPE NETMASK, set to 0x00, as specified by this document for all requests.

    *ADDRESS, set to 0xC0 0x00 0x02, providing only the first 24 bits of the IPv4 address.

6. The query is sent. Authoritative Nameserver ANS understands and uses edns-client-subnet. It parses the edns-client-subnet option, and generates an optimized reply.

7. Due to the internal implementation of the Authoritative Nameserver ANS, ANS finds a reply that is optimal for the whole /16 of the client that performed the request.

8. The Authoritative Nameserver ANS adds an edns-client-subnet option in the reply, containing:

    *OPTION-CODE, set to TBD.

    *OPTION-LENGTH, set to 0x00 0x07.

    *FAMILY, set to 0x00 0x01.

    *SOURCE NETMASK, set to 0x18, copied from the request.

    *SCOPE NETMASK, set to 0x10, indicating a /16 network range.

    *ADDRESS, set to 0xC0 0x00 0x02, copied from the request.

9. The Recursive Resolver R receives the reply containing an edns-client-subnet option. The resolver verifies that FAMILY, SOURCE NETMASK, and ADDRESS match the request. If not, the option is discarded.

10. The reply is interpreted as usual. Since the reply contains an edns-client-subnet option, the ADDRESS, SCOPE NETMASK, and FAMILY in the response are used to cache the entry.

11. R sends a response to stub resolver SR, without including an edns-client-subnet option.

12. R receives another request to resolve www.example.com. This time, a reply is cached. The reply, however, is tied to a particular network. If the address of the client matches any network in the cache, then the reply is returned from the cache. Otherwise, another query is performed. If multiple results match, the one with the longest SCOPE NETMASK is chosen, as per common best-network match algorithms.

---

10.   Acknowledgements

Appendix A.   Document Editing History

Appendix A.1.   Changes since -00

   *Rewritten problem statement to be more clear about the goal of edns-client-subnet and the fact that it's entirely optional.

   *Wire format changed to include the original address and netmask in responses in defence against birthday attacks.

   *Security considerations now includes a section about birthday attacks.

   *Renamed edns-client-ip in edns-client-subnet, following suggestions on the mailing list.

*Clarified behavior of resolvers when presented with an invalid
  edns-client-subnet option.

*Fully take multi-tier DNS setups in mind and be more clear about
  where the option should be originated.

*Added a few definitions in the Terminology section, and a few
  more aesthetic changes in the rest of the document.

---

11. References

11.1. Normative References

[RFC1034]	Mockapetris, P., "Domain names - concepts and facilities," STD 13, RFC 1034, November 1987 (TXT).
[RFC1035]	Mockapetris, P., "Domain names - implementation and specification," STD 13, RFC 1035, November 1987 (TXT).
[RFC1918]	Rekhter, Y., Moskowitz, R., Karrenberg, D., Groot, G., and E. Lear, "Address Allocation for Private Internets," BCP 5, RFC 1918, February 1996 (TXT).
[RFC2119]	Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels," BCP 14, RFC 2119, March 1997 (TXT, HTML, XML).
[RFC2671]	Vixie, P., "Extension Mechanisms for DNS (EDNS0)," RFC 2671, August 1999 (TXT).
[RFC4033]	Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "DNS Security Introduction and Requirements," RFC 4033, March 2005 (TXT).
[RFC4034]	Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "Resource Records for the DNS Security Extensions," RFC 4034, March 2005 (TXT).
[RFC4035]	Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "Protocol Modifications for the DNS Security Extensions," RFC 4035, March 2005 (TXT).
[RFC4193]	Hinden, R. and B. Haberman, "Unique Local IPv6 Unicast Addresses," RFC 4193, October 2005 (TXT).

11.2. Informative References

[RFC2663]

Srisuresh, P. and M. Holdrege, "IP Network Address Translator (NAT) Terminology and Considerations," RFC 2663.

Authors' Addresses

TOC

	Carlo Contavalli
	Google
	Gordon House, Barrow Street
	Dublin 4
	IE
Email:	ccontavalli@google.com
	Wilmer van der Gaast
	Google
	Gordon House, Barrow Street
	Dublin 4
	IE
Email:	wilmer@google.com
	Sean Leach
	Name.com
	125 Rampart Way, Suite 300
	Denver, CO 80230
	CO
Email:	sleach@name.com
	Darryl Rodden
	Neustar
	46000 Center Oak Plaza
	Sterling, VA 20166
	VA
Email:	darryl.rodden@neustar.com