dprive
Internet-Draft
Obsoletes: 7626 (if approved)
Intended status: Informational
Expires: March 30, 2020

S. Bortzmeyer
AFNIC
S. Dickinson
Sinodun IT

September 27, 2019

DNS Privacy Considerations
draft-ietf-dprive-rfc7626-bis-01

Abstract

This document ~~describes~~ identifies the privacy issues associated with the use of the DNS by Internet users. It is intended to be an analysis of the present situation and does not prescribe solutions. This document obsoletes RFC 7626.

Table of Contents

1. Introduction

This document is an analysis of the DNS privacy issues, in the spirit of Section 8 of [RFC6973].

The Domain Name System (DNS) is specified in [RFC1034], [RFC1035], and many later RFCs, which have never been consolidated. It is one of the most important infrastructure components of the Internet and often ignored or misunderstood by Internet users (and even by many professionals). Almost every activity on the Internet starts with a DNS query (and often several). Its use has many privacy implications and this document is an attempt at a comprehensive and accurate list.

Commentaire [Med1]: I'm not sure I would maintain this.

Let us begin with a simplified reminder of how the DNS works. (See also [RFC8499]) A client, the stub resolver, issues a DNS query to a server, called the recursive resolver (also called caching resolver or full resolver or recursive name server). Let's use the query "What are the AAAA records for www.example.com?" as an example. AAAA is the QTYPE (Query Type), and www.example.com is the QNAME (Query Name). (The description that follows assumes a cold cache, for

instance, because the server just started.) The recursive resolver will first query the root name servers. In most cases, the root name servers will send a referral. In this example, the referral will be to the .com name servers. The resolver repeats the query to one of the .com name servers. The .com name servers, in turn, will refer to the example.com name servers. The example.com name server will then return the answer. The root name servers, the name servers of .com, and the name servers of example.com are called authoritative name servers. It is important, when analyzing the privacy issues, to remember that the question asked to all these name servers is always the original question, not a derived question. The question sent to the root name servers is "What are the AAAA records for www.example.com?", not "What are the name servers of .com?". By repeating the full question, instead of just the relevant part of the question to the next in line, the DNS provides more information than necessary to the name server. In this simplified description, recursive resolvers do not implement QNAME minimization as described in [RFC7816], which will only send the relevant part of the question to the upstream name server.

Because DNS relies on caching heavily, the algorithm described ~~just~~ above is actually a bit more complicated, and not all questions are sent to the authoritative name servers. If a few seconds later the stub resolver asks the recursive resolver, "What are the SRV records of _xmpp-server._tcp.example.com?", the recursive resolver will remember that it knows the name servers of example.com and will just query them, bypassing the root and .com. Because there is typically no caching in the stub resolver, the recursive resolver, unlike the authoritative servers, sees all the DNS traffic. (Applications, like web browsers, may have some form of caching that does not follow DNS rules, for instance, because it may ignore the TTL. So, the recursive resolver does not see all the name resolution activity.)

It should be noted that DNS recursive resolvers sometimes forward requests to other recursive resolvers, typically 'bigger' machines, with a larger and more shared cache (and the query hierarchy can be even deeper, with more than two levels of recursive resolvers). From the point of view of privacy, these forwarders are like resolvers, except that they do not see all of the requests being made (due to caching in the first resolver).

At the time of writing (2019), almost all this DNS traffic is currently sent in clear (i.e., unencrypted). However there is increasing deployment of DNS-over-TLS (DoT) [RFC7858] and DNS-over-HTTPS (DoH) [RFC8484], particularly in mobile devices, browsers, and by providers of anycast recursive DNS resolution services. There are a few cases where there is some alternative channel encryption, for instance, in an IPsec ~~VPN~~tunnel, at least between the stub resolver and the resolver.

Commentaire [Med2]: I see that you listed this one as normative, while it should be listed as informative.

Today, almost all DNS queries are sent over UDP [thomas-ditl-tcp]. This has practical consequences when considering encryption of the traffic as a possible privacy technique. Some encryption solutions are only designed for TCP, not UDP and new solutions are still emerging [I-D.ietf-quic-transport].

Commentaire [Med3]: I guess you exclude RFC8094

Another important point to keep in mind when analyzing the privacy issues of DNS is ~~the fact~~ that DNS requests received by a server are triggered by different reasons. For example, ~~let's~~ let's assume an eavesdropper wants to know which web page is viewed by a user. For a typical web page, there are three sorts of DNS requests being issued:

- Primary request: this is the domain name in the URL that the user typed, selected from a bookmark, or chose by clicking on an hyperlink. Presumably, this is what is of interest for the eavesdropper.

- Secondary requests: these are the additional requests performed by the user agent (here, the web browser) without any direct involvement or knowledge of the user. For the Web, they are triggered by embedded content, Cascading Style Sheets (CSS), JavaScript code, embedded images, etc. In some cases, there can be dozens of domain names in different contexts on a single web page.

- Tertiary requests: these are the additional requests performed by the DNS system itself. For instance, if the answer to a query is a referral to a set of name servers, and the glue records are not returned, the resolver will have to do additional requests to turn the name servers' names into IP addresses. Similarly, even if glue records are returned, a careful recursive server will do tertiary requests to verify the IP addresses of those records.

It can be noted also that, in the case of a typical web browser, more DNS requests than strictly necessary are sent, for instance, to prefetch resources that the user may query later or when autocompleting the URL in the address bar. Both are a big privacy concern since they may leak information even about non-explicit actions. For instance, just reading a local HTML page, even without selecting the hyperlinks, may trigger DNS requests.

For privacy-related terms, we will use the terminology from [RFC6973].

2. Scope

This document focuses mostly on the study of privacy risks for the end user (the one ~~performing~~ sending DNS requests). We consider the risks of pervasive surveillance [RFC7258] as well as risks coming from a more focused surveillance.

Privacy risks for the holder of a zone (the risk that someone gets the data) are discussed in [RFC5936] and [RFC5155].

Privacy risks for recursive operators such as leakage of private namespaces or blocklists are out of scope for this document.

Non-privacy risks (e.g., security related concerns such as cache poisoning) are also out of scope.

DNS may also be used within closed environments (e.g., Enterprise networks). The document does not elaborate on any specifics related to those deployments.

Despite the document distinguishes resolvers operated by access providers and public ones, there is no evidence about a common privacy profile followed by all access providers (or public resolvers). Also, it is not the intent of this document to make a comparison between privacy protections provided by access providers vs. public resolvers.

3. Risks

3.1. The Alleged Public Nature of DNS Data

It has long been claimed that "the data in the DNS is public". While this sentence makes sense for an Internet-wide lookup system, there are multiple facets to the data and metadata involved that deserve a more detailed look. First, access control lists (ACLs) and private namespaces notwithstanding, the DNS operates under the assumption that public-facing authoritative name servers will respond to "usual" DNS queries for any zone they are authoritative for without further authentication or authorization of the client (resolver). Due to the lack of search capabilities, only a given QNAME will reveal the resource records associated with that name (or that name's non-existence). In other words, ~~:~~ one needs to know what to ask for, in order to receive a response. The zone transfer QTYPE [RFC5936] is often blocked or restricted to authenticated/authorized access to enforce this difference (and maybe for other reasons).

Another differentiation to be considered is between the DNS data itself and a particular transaction (i.e., a DNS name lookup). DNS data and the results of a DNS query are public, within the boundaries described above, and may not have any confidentiality requirements. However, the same is not true of a single transaction or a sequence of transactions; that transaction is not / should not be public. A typical example from outside the DNS world is: the web site of Alcoholics Anonymous is public; the fact that you visit it should not

be.

3.2. Data in the DNS Request

The DNS request includes many fields, but two of them seem particularly relevant for the privacy issues: the QNAME and the source IP address. "source IP address" is used in a loose sense of "source IP address + maybe source port number", because the port number is also in the request and can be used to differentiate between several users sharing an IP address (behind a Carrier-Grade NAT (CGN) or a NPTv6, for instance [RFC6269]).

The QNAME is the full name sent by the user. It gives information about what the user does ("What are the MX records of example.net?" means he probably wants to send an email to someone at example.net, which may be a domain used by only a few persons and is therefore very revealing about communication relationships). Some QNAMEs are more sensitive than others. For instance, querying the A record of a well-known web statistics domain reveals very little (everybody visits web sites that use this analytics service), but querying the A record of www.verybad.example where verybad.example is the domain of an organization that some people find offensive or objectionable may create more problems for the user. Also, sometimes, the QNAME embeds the software one uses, which could be a privacy issue. For instance, _ldap._tcp.Default-First-Site-Name._sites.gc._msdcs.example.org. There are also some BitTorrent clients that query an SRV record for _bittorrent-tracker._tcp.domain.example.

Another important thing about the privacy of the QNAME is the future usages. Today, the lack of privacy is an obstacle to putting potentially sensitive or personally identifiable data in the DNS. At the moment, your DNS traffic might reveal that you are doing email but not with whom. If your Mail User Agent (MUA) starts looking up Pretty Good Privacy (PGP) keys in the DNS [RFC7929], then privacy becomes a lot more important. And email is just an example; there would be other really interesting uses for a more privacy-friendly~~privacy - friendly~~ DNS.

For the communication between the stub resolver and the recursive resolver, the source IP address is the address of the user's machine. Therefore, all the issues and warnings about collection of IP addresses apply here. For the communication between the recursive resolver and the authoritative name servers, the source IP address has a different meaning; it does not have the same status as the source address in an HTTP connection. It is now the IP address of the recursive resolver that, in a way, "hides" the real user. However, hiding does not always work. Sometimes EDNS(0) Client subnet [RFC7871] is used (see its privacy analysis in [denis-edns-client-subnet]). Sometimes the end user has a personal

Commentaire [Med4]: To cover the v6 case.

Commentaire [Med5]: This may not be that accurate if the host is behind a CPE (with the recursive resolver). Both belong to the user.

recursive resolver on her machine. In both cases, the IP address is as sensitive as it is for HTTP [sidn-entrada].

A note about IP addresses: there is currently no IETF document that describes in detail all the privacy issues around IP addressing. In the meantime, the discussion here is intended to include both IPv4 and IPv6 source addresses. For a number of reasons, their assignment and utilization characteristics are different, which may have implications for details of information leakage associated with the collection of source addresses. (For example, a specific IPv6 source address seen on the public Internet is less likely than an IPv4 address to originate behind ~~a CGN or other NAT~~an address sharing scheme.) However, for both IPv4 and IPv6 addresses, it's important to note that source addresses are propagated with queries and comprise metadata about the host, user, or application that originated them.

3.2.1. Data in the DNS payload

At the time of writing there are no standardized client identifiers contained in the DNS payload itself (ECS [RFC7871] while widely used is only of Category Informational).

DNS Cookies [RFC7873] are a lightweight DNS transaction security mechanism that provides limited protection against a variety of increasingly common denial-of-service and amplification/forgery or cache poisoning attacks by off-path attackers. It is noted, however, that they are designed to ~~just~~ verify IP addresses (and should change once a client's IP address changes), they are not designed to actively track users (like HTTP cookies).

There are anecdotal accounts of MAC addresses [1] and even user names being inserted in non-standard EDNS(0) options for stub to resolver communications to support proprietary functionality implemented at the resolver (e.g., parental filtering).

3.3. Cache Snooping

The content of recursive resolvers' caches can reveal data about the clients using it (the privacy risks depend on the number of clients). This information can sometimes be examined by sending DNS queries with RD=0 to inspect cache content, particularly looking at the DNS TTLs [grangeia.snooping]. Since this also is a reconnaissance technique for subsequent cache poisoning attacks, some counter measures have already been developed and deployed.

3.4. On the Wire

3.4.1. Unencrypted Transports

For unencrypted transports, DNS traffic can be seen by an eavesdropper like any other traffic. (DNSSEC, specified in [RFC4033], explicitly excludes confidentiality from its goals.) So, if an initiator starts an HTTPS communication with a recipient, while the HTTP traffic will be encrypted, the DNS exchange prior to it will not be. When other protocols will become more and more privacy-aware and secured against surveillance (e.g., [RFC8446], [I-D.ietf-quic-transport]), the use of unencrypted transports for DNS may become "the weakest link" in privacy. It is noted that at the time of writing there is on-going work attempting to encrypt the SNI in the TLS handshake [I-D.ietf-tls-sni-encryption].

An important specificity of the DNS traffic is that it may take a different path than the communication between the initiator and the recipient. For instance, an eavesdropper may be unable to tap the wire between the initiator and the recipient but may have access to the wire going to the recursive resolver, or to the authoritative name servers.

The best place to tap, from an eavesdropper's point of view, is clearly between the stub resolvers and the recursive resolvers, because traffic is not limited by DNS caching.

The attack surface between the stub resolver and the rest of the world can vary widely depending upon how the end user's computer is configured. By order of increasing attack surface:

The recursive resolver can be on the end user's computer. In (currently) a small number of cases, individuals may choose to operate their own DNS resolver on their local machine. In this case, the attack surface for the connection between the stub resolver and the caching resolver is limited to that single machine.

The recursive resolver may be at the local network edge. For many/most enterprise networks and for some residential users, the caching resolver may exist on a server at the edge of the local network. In this case, the attack surface is the local network. Note that in large enterprise networks, the DNS resolver may not be located at the edge of the local network but rather at the edge of the overall enterprise network. In this case, the enterprise network could be thought of as similar to the Internet Access Provider (IAP) network referenced below.

The recursive resolver can be in the IAP premises. For most residential users and potentially other networks, the typical case is for the end user's computer to be configured (typically automatically through DHCP or PCO for cellular networks) with the addresses of the DNS recursive resolvers at the IAP. The attack surface for on-the-wire attacks is therefore from the end-user system across the local network and across the IAP network to the IAP's recursive resolvers.

The recursive resolver can be a public DNS service. Some machines may be configured to use public DNS resolvers such as those operated today by Google Public DNS or OpenDNS. The end user may have configured their machine to use these DNS recursive resolvers themselves -- or their IAP may have chosen to use the public DNS resolvers rather than operating their own resolvers. In this case, the attack surface is the entire public Internet between the end user's connection and the public DNS service.

3.4.2. Encrypted Transports

The use of encrypted transports directly mitigates passive surveillance of the DNS payload, however there are still some privacy attacks possible. This section enumerates the residual privacy risks to an end user when an attacker can passively monitor encrypted DNS traffic flows on the wire.

These are cases where user identification, fingerprinting or correlations may be possible due to the use of certain transport layers or clear text/observable features. These issues are not specific to DNS, but DNS traffic is susceptible to these attacks when using specific transports.

Commentaire [Med6]: Still there is possibility to correlate the destination IP address with a name. See for example, https://dl.acm.org/citation.cfm?id=3341133

There are some general examples, for example, certain studies have highlighted that IP TTL or TCP Window sizes os-fingerprint [2] values can be used to fingerprint client OS's or that various techniques can be used to de-NAT DNS queries dns-de-nat [3].

The use of clear text transport options to decrease latency may also identify a user, e.g., using TCP Fast Open [RFC7413].

Commentaire [Med7]: May be useful to elaborate more.

More specifically, (since the deployment of encrypted transports is not widespread at the time of writing) users wishing to use encrypted transports for DNS may in practice be limited in the resolver services available. Given this, the choice of a user to configure a single resolver (or a fixed set of resolvers) and an encrypted transport to use in all network environments can actually serve to identify the user as one that desires privacy and can provide an added mechanism to track them as they move across network environments.

Users of encrypted transports are also highly likely to re-use sessions for multiple DNS queries to optimize performance (e.g., via DNS pipelining or HTTPS multiplexing). Certain configuration options for encrypted transports could also in principle fingerprint a user or client application. For example:

- TLS version or cipher suite selection
- session resumption
- the maximum number of messages to send or
- a maximum connection time before closing a connections and re-opening.

Whilst there are known attacks on older versions of TLS the most recent recommendations [RFC7525] and developments [RFC8446] in this area largely mitigate those.

Traffic analysis of unpadded encrypted traffic is also possible [pitfalls-of-dns-encrption] because the sizes and timing of encrypted DNS requests and responses can be correlated to unencrypted DNS requests upstream of a recursive resolver.

3.5. In the Servers

Using the terminology of [RFC6973], the DNS servers (recursive resolvers and authoritative servers) are enablers: they facilitate communication between an initiator and a recipient without being directly in the communications path. As a result, they are often

forgotten in risk analysis. But, to quote again [RFC6973], "Although [...] enablers may not generally be considered as attackers, they may all pose privacy threats (depending on the context) because they are able to observe, collect, process, and transfer privacy-relevant data." In [RFC6973] parlance, enablers become observers when they start collecting data.

Many programs exist to collect and analyze DNS data at the servers -- from the "query log" of some programs like BIND to tcpdump and more sophisticated programs like PacketQ [packetq] and DNSmezzo [dnsmezzo]. The organization managing the DNS server can use this data itself, or it can be part of a surveillance program like PRISM [prism] and pass data to an outside observer.

Sometimes, this data is kept for a long time and/or distributed to third parties for research purposes [ditl] [day-at-root], security analysis, or surveillance tasks. These uses are sometimes under some sort of contract, with various limitations, for instance, on redistribution, given the sensitive nature of the data. Also, there are observation points in the network that gather DNS data and then make it accessible to third parties for research or security purposes ("passive DNS" [passive-dns]).

3.5.1. In the Recursive Resolvers

Recursive Resolvers see all the traffic since there is typically no caching before them. To summarize: your recursive resolver knows a lot about you. The resolver of a large IAP, or a large public resolver, can collect data from many users.

3.5.1.1. Resolver ~~selection~~Selection

Given all the above considerations the choice of recursive resolver has direct privacy considerations for end users. Historically end user devices have used the DHCP (or other specific means such as PCO in cellular networks) provided local network recursive resolver which may have strong, medium or weak privacy policies depending on the network. Privacy policies for these servers may or may not be available and users need to be aware that privacy guarantees will vary with network.

More recently some networks and end users have actively chosen to use a large public resolver instead, e.g., Google Public DNS, Cloudflare or Quad9 (need refs). There can be many reasons: cost considerations for network operators, better reliability or anti-censorship considerations are just a few. Such services typically do provide a privacy policy and the end user can get an idea of the data collected by such operators by reading one, e.g., Google Public DNS - Your Privacy [4].

Even more recently some applications have announced plans to deploy ~~application~~ application-specific DNS settings which might be enabled by default.
For example current proposals by Firefox [firefox] revolve around a default based on geographic region using a pre-configured list of large public resolver services which offer DoH, combined with non-

standard probing and signalling mechanism to disable DoH in

particular networks. Whereas Chrome [chrome] is experimenting with

using DoH to the DHCP provided resolver if it is on a list of DoH-

compatible providers. At the time of writing efforts to provide

standardized signalling mechanisms for applications to discover the services offered by local resolvers are in progress [I-D.ietf-dnsop-resolver-information].

If applications enable ~~application~~ application-specific DNS settings without
properly informing the user of the change (or do not provide an option for user configuration of the application recursive resolver) there is a potential privacy issue; depending on the network context and the application default the application might use a recursive server that provides less privacy protection than the default network provided server without the users full knowledge. Users that are fully aware of an application specific DNS setting may want to actively override any default in favour of their chosen recursive resolver.

There are also concerns that should the trend towards using large public resolvers increase, this will itself provide a privacy concern due to a small number of operators having visibility of the majority of DNS requests globally and the potential for aggregating data across services about a user. Additionally the operating organisation of the resolver may be in a different legal jurisdiction to the user which creates further privacy concerns around legal protections of and access to the data collected by the operator.

At the time of writing the deployment models for DNS are evolving, their implications are complex and extend beyond the scope of this document. They are the subject of much other work including [I-D.livingood-doh-implementation-risks-issues], the IETF ADD mailing list [5] and the Encrypted DNS Deployment Initiative [6].

3.5.1.2. Active ~~attacks~~ Attacks on ~~resolver~~ Resolver ~~configuration~~Configuration

The previous ~~paragraphs~~ section discussed DNS privacy, assuming that all the
traffic was directed to the intended servers (i.e., those that would be
used in the absence of an active attack) and that the potential

attacker was purely passive. But, in reality, we can have active attackers in the network redirecting the traffic, not just to observe it but also potentially change it.

For instance, a DHCP server controlled by an attacker can direct you to a recursive resolver also controlled by that attacker. Most of the time, it seems to be done to divert traffic in order to also direct the user to a web server controlled by the attacker. However it could be used just to capture the traffic and gather information about you.

Other attacks, besides using DHCP, are possible. The cleartext traffic from a DNS client to a DNS server can be intercepted along its way from originator to intended source, for instance, by transparent attacker controlled DNS proxies in the network that will divert the traffic intended for a legitimate DNS server. This server can masquerade as the intended server and respond with data to the client. (Attacker controlled servers that inject malicious data are possible, but it is a separate problem not relevant to privacy.) A server controlled by an attacker may respond correctly for a long period of time, thereby foregoing detection.

Also, malware like DNSchanger [dnschanger] can change the recursive resolver in the machine's configuration, or the routing itself can be subverted (for instance, [ripe-atlas-turkey]).

3.5.1.3. Blocking of ~~user~~ User ~~selected~~ Selected ~~services~~Services

User privacy can also be at risk if there is blocking (by local network operators or more general mechanisms) of access to remote recursive servers that offer encrypted transports when the local resolver does not offer encryption and/or has very poor privacy policies. For example, active blocking of port 853 for DoT or of specific IP addresses (e.g., 1.1.1.1 or 2606:4700:4700::1111) could restrict the resolvers available to the user. The extent of the risk to end user privacy is highly ~~dependant~~dependent on the specific network and user context; a user on a network that is known to perform surveillance would be compromised if they could not access such services whereas a user on a trusted network might have no privacy motivation to do so.

Similarly attacks on such services, e.g., DDoS could force users to switch to other services that do not offer encrypted transports for DNS.

3.5.1.4. Authentication of Servers

Both DoH and "Strict mode" for DoT require authentication of the server and therefore as long as the authentication credentials are obtained over a secure channel then using either of these transports defeats the attack of re-directing traffic to rogue servers. Of course

Commentaire [Med8]: Some networks, e.g., enterprises, can enforce similar filtering for specific reasons that are not related to privacy.

Commentaire [Med9]: Please add a pointer to RFC8310

attacks on these secure channels are also possible, but out of the scope of this document.

3.5.1.5. Encrypted Transports

3.5.1.5.1. DoT and DoH

Use of encrypted transports does not reduce the data available in the recursive resolver and ironically can actually expose more information about users to operators. As mentioned in Section 3.4 use of session based encrypted transports (TCP/TLS) can expose correlation data about users. Such concerns in the TCP/TLS layers apply equally to DoT and DoH which both use TLS as the underlying transport, some examples are:

- fingerprinting based on TLS version and/or cipher suite selection
- user tracking via session resumption in TLS 1.2

3.5.1.5.2. DoH Specific Considerations

The proposed specification for DoH [RFC8484] includes a 'Privacy Considerations' section which highlights some of the differences between HTTP and DNS. As a deliberate design choice DoH inherits the privacy properties of the HTTPS stack and as a consequence introduces new privacy concerns when compared with DNS over UDP, TCP or TLS [RFC7858]. The rationale for this decision is that retaining the ability to leverage the full functionality of the HTTP ecosystem is more important than placing specific constraints on this new protocol based on privacy considerations (modulo limiting the use of HTTP cookies).

In analyzing the new issues introduced by DoH it is helpful to recognize that there exists a natural tension between

- the wide practice in HTTP to use various headers to optimize HTTP connections, functionality and behaviour (which can facilitate user identification and tracking)
- and the fact that the DNS payload is currently very tightly encoded and contains no standardized user identifiers.

DoT, for example, would normally contain no client identifiers above the TLS layer and a resolver would see only a stream of DNS query payloads originating within one or more connections from a client IP address. Whereas if DoH clients commonly include several headers in a DNS message (e.g., user-agent and accept-language) this could lead to the DoH server being able to identify the source of individual DNS

requests not only to a specific end user device but to a specific application.

Additionally, depending on the client architecture, isolation of DoH queries from other HTTP traffic may or may not be feasible or desirable. Depending on the use case, isolation of DoH queries from other HTTP traffic may or may not increase privacy.

The picture for privacy considerations and user expectations for DoH with respect to what additional data may be available to the DoH server compared to DNS over UDP, TCP or TLS is complex and requires a detailed analysis for each use case. In particular the choice of HTTPS functionality vs privacy is specifically made an implementation choice in DoH and users may well have differing privacy expectations depending on the DoH use case and implementation.

At the extremes, there may be implementations that attempt to achieve parity with DoT from a privacy perspective at the cost of using no identifiable headers, there might be others that provide feature rich data flows where the low-level origin of the DNS query is easily identifiable.

Privacy ~~focussed~~focused users should be aware of the potential for additional client identifiers in DoH compared to DoT and may want to only use DoH client implementations that provide clear guidance on what identifiers they add.

3.5.2. In the Authoritative Name Servers

Unlike what happens for recursive resolvers, observation capabilities of authoritative name servers are limited by caching; they see only the requests for which the answer was not in the cache. For aggregated statistics ("What is the percentage of LOC queries?"), this is sufficient, but it prevents an observer from seeing everything. Similarly the increasing deployment of QNAME minimisation [ripe-qname-measurements] reduces the data visible at the authoritative name server. Still, the authoritative name servers see a part of the traffic, and this subset may be sufficient to violate some privacy expectations.

Also, the end user typically has some legal/contractual link with the recursive resolver (he has chosen the IAP, or he has chosen to use a given public resolver), while having no control and perhaps no awareness of the role of the authoritative name servers and their observation abilities.

As noted before, using a local resolver or a resolver close to the machine decreases the attack surface for an on-the-wire eavesdropper.

But it may decrease privacy against an observer located on an authoritative name server. This authoritative name server will see the IP address of the end client instead of the address of a big recursive resolver shared by many users.

This "protection", when using a large resolver with many clients, is no longer present if ECS [RFC7871] is used because, in this case, the authoritative name server sees the original IP address (or prefix, depending on the setup).

As of today, all the instances of one root name server, L-root, receive together around 50,000 queries per second. While most of it is "junk" (errors on the Top-Level Domain (TLD) name), it gives an idea of the amount of big data that pours into name servers. (And even "junk" can leak information; for instance, if there is a typing error in the TLD, the user will send data to a TLD that is not the usual one.)

Many domains, including TLDs, are partially hosted by third-party servers, sometimes in a different country. The contracts between the domain manager and these servers may or may not take privacy into account. Whatever the contract, the third-party hoster may be honest or not but, in any case, it will have to follow its local laws. So, requests to a given ccTLD may go to servers managed by organizations outside of the ccTLD's country. End users may not anticipate that, when doing a security analysis.

Also, it seems (see the survey described in [aeris-dns]) that there is a strong concentration of authoritative name servers among "popular" domains (such as the Alexa Top N list). For instance, among the Alexa Top 100K, one DNS provider hosts today 10% of the domains. The ten most important DNS providers host together one third of the domains. With the control (or the ability to sniff the traffic) of a few name servers, you can gather a lot of information.

3.6. Re-identification and Other Inferences

An observer has access not only to the data he/she directly collects but also to the results of various inferences about this data. The term 'observer' here is used very generally, it might be one that is passively observing cleartext DNS traffic, one in the network that is actively attacking the user by re-directing DNS resolution, or it might be a local or remote resolver operator.

For instance, a user can be re-identified via DNS queries. If the adversary knows a user's identity and can watch their DNS queries for a period, then that same adversary may be able to re-identify the user solely based on their pattern of DNS queries later on regardless

of the location from which the user makes those queries. For example, one study [herrmann-reidentification] found that such re-identification is possible so that "73.1% of all day-to-day links were correctly established, i.e., user u was either re-identified unambiguously (1) or the classifier correctly reported that u was not present on day t+1 any more (2)." While that study related to web browsing behavior, equally characteristic patterns may be produced even in machine-to-machine communications or without a user taking specific actions, e.g., at reboot time if a characteristic set of services are accessed by the device.

For instance, one could imagine that an intelligence agency identifies people going to a site by putting in a very long DNS name and looking for queries of a specific length. Such traffic analysis could weaken some privacy solutions.

The IAB privacy and security program also have a work in progress [RFC7624] that considers such inference-based attacks in a more general framework.

3.7. More Information

Useful background information can also be found in [tor-leak] (about the risk of privacy leak through DNS) and in a few academic papers: [yanbin-tsudik], [castillo-garcia], [fangming-hori-sakurai], and [federrath-fuchs-herrmann-piosecny].

4. Actual "Attacks"

A very quick examination of DNS traffic may lead to the false conclusion that extracting the needle from the haystack is difficult. "Interesting" primary DNS requests are mixed with useless (for the eavesdropper) secondary and tertiary requests (see the terminology in Section 1). But, in this time of "big data" processing, powerful techniques now exist to get from the raw data to what the eavesdropper is actually interested in.

Many research papers about malware detection use DNS traffic to detect "abnormal" behavior that can be traced back to the activity of malware on infected machines. Yes, this research was done for the good, but technically it is a privacy attack and it demonstrates the power of the observation of DNS traffic. See [dns-footprint], [dagon-malware], and [darkreading-dns].

Passive DNS systems [passive-dns] allow reconstruction of the data of sometimes an entire zone. They are used for many reasons -- some good, some bad. Well-known passive DNS systems keep only the DNS responses, and not the source IP address of the client, precisely for

privacy reasons. Other passive DNS systems may not be so careful. And there is still the potential problems with revealing QNAMEs.

The revelations from the Edward Snowden documents, which were leaked from the National Security Agency (NSA) provide evidence of the use of the DNS in mass surveillance operations [morecowbell]. For example the MORECOWBELL surveillance program, which uses a dedicated covert monitoring infrastructure to actively query DNS servers and perform HTTP requests to obtain meta information about services and to check their availability. Also the QUANTUMTHEORY project which includes detecting lookups for certain addresses and injecting bogus replies is another good example showing that the lack of privacy protections in the DNS is actively exploited.

5. Legalities

To our knowledge, there are no specific privacy laws for DNS data, in any country. Interpreting general privacy laws like [data-protection-directive] or GDPR [7] applicable in the European Union in the context of DNS traffic data is not an easy task, and we do not know a court precedent here. See an interesting analysis in [sidn-entrada].

6. Security Considerations

This document is entirely about security, more precisely privacy. It just lays out the problem; it does not try to set requirements (with the choices and compromises they imply), much less define solutions. Possible solutions to the issues described here are discussed in other documents (currently too many to all be mentioned); see, for instance, 'Recommendations for DNS Privacy Operators' [I-D.ietf-dprive-bcp-op].

7. Acknowledgments

Thanks to Nathalie Boulvard and to the CENTR members for the original work that led to this document. Thanks to Ondrej Sury for the interesting discussions. Thanks to Mohsen Souissi and John Heidemann for proofreading and to Paul Hoffman, Matthijs Mekking, Marcos Sanz, Tim Wicinski, Francis Dupont, Allison Mankin, and Warren Kumari for proofreading, providing technical remarks, and making many readability improvements. Thanks to Dan York, Suzanne Woolf, Tony Finch, Stephen Farrell, Peter Koch, Simon Josefsson, and Frank Denis for good written contributions. And thanks to the IESG members for the last remarks.

8. Changelog

draft-ietf-dprive-rfc7627-bis-01

- Re-structure section 3.5 (was 2.5)
  - Collect considerations for recursive resolvers together
  - Re-work several sections here to clarify their context (e.g. 'Rogue servers' becomes 'Active attacks on resolver configuration')
  - Add discussion of resolver selection
- Update text and old reference on Snowdon revelations.
- Add text on and references to QNAME minimisation RFC and deployment measurements
- Correct outdated references
- Clarify scope by adding a Scope section (was Risks overview)
- Clarify what risks are considered in section 3.4.2

draft-ietf-dprive-rfc7627-bis-00

- Rename after WG adoption
- Use DoT acronym throughout
- Minor updates to status of deployment and other drafts

draft-bortzmeyer-dprive-rfc7626-bis-02

- Update various references and fix some nits.

draft-bortzmeyer-dprive-rfc7626-bis-01

- Update reference for dickinson-bcp-op to draft-dickinson-dprive-bcp-op

draft-borztmeyer-dprive-rfc7626-bis-00:

Initial commit. Differences to RFC7626:

- Update many references

- Add discussions of encrypted transports including DoT and DoH
- Add section on DNS payload
- Add section on authentication of servers
- Add section on blocking of services

9. References

9.1. Normative References

[RFC1034] Mockapetris, P., "Domain names - concepts and facilities", STD 13, RFC 1034, DOI 10.17487/RFC1034, November 1987, <https://www.rfc-editor.org/info/rfc1034>.

[RFC1035] Mockapetris, P., "Domain names - implementation and specification", STD 13, RFC 1035, DOI 10.17487/RFC1035, November 1987, <https://www.rfc-editor.org/info/rfc1035>.

[RFC6973] Cooper, A., Tschofenig, H., Aboba, B., Peterson, J., Morris, J., Hansen, M., and R. Smith, "Privacy Considerations for Internet Protocols", RFC 6973, DOI 10.17487/RFC6973, July 2013, <https://www.rfc-editor.org/info/rfc6973>.

[RFC7258] Farrell, S. and H. Tschofenig, "Pervasive Monitoring Is an Attack", BCP 188, RFC 7258, DOI 10.17487/RFC7258, May 2014, <https://www.rfc-editor.org/info/rfc7258>.

[RFC7816] Bortzmeyer, S., "DNS Query Name Minimisation to Improve Privacy", RFC 7816, DOI 10.17487/RFC7816, March 2016, <https://www.rfc-editor.org/info/rfc7816>.

Commentaire [Med10]: Not sure why it is listed as normative

9.2. Informative References

[aeris-dns] Vinot, N., "Vie privee: et le DNS alors?", (In French), 2015, <https://blog.imirhil.fr/vie-privee-et-le-dns-alors.html>.

[castillo-garcia] Castillo-Perez, S. and J. Garcia-Alfaro, "Anonymous Resolution of DNS Queries", 2008, <http://deic.uab.es/~joaquin/papers/is08.pdf>.

[chrome] Baheux, , "Experimenting with same-provider DNS-over-HTTPS upgrade", September 2019, <https://blog.chromium.org/2019/09/experimenting-with-same-provider-dns.html>.

[dagon-malware] Dagon, D., "Corrupted DNS Resolution Paths: The Rise of a Malicious Resolution Authority", ISC/OARC Workshop, 2007, <https://www.dns-oarc.net/files/workshop-2007/Dagon-Resolution-corruption.pdf>.

[darkreading-dns] Lemos, R., "Got Malware? Three Signs Revealed In DNS Traffic", InformationWeek Dark Reading, May 2013, <http://www.darkreading.com/analytics/security-monitoring/got-malware-three-signs-revealed-in-dns-traffic/d/d-id/1139680>.

[data-protection-directive] European Parliament, "Directive 95/46/EC of the European Pariament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data", Official Journal L 281, pp. 0031 - 0050, November 1995, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML>.

[day-at-root] Castro, S., Wessels, D., Fomenkov, M., and K. Claffy, "A Day at the Root of the Internet", ACM SIGCOMM Computer Communication Review, Vol. 38, Number 5, DOI 10.1145/1452335.1452341, October 2008, <http://www.sigcomm.org/sites/default/files/ccr/papers/2008/October/1452335-1452341.pdf>.

[denis-edns-client-subnet] Denis, F., "Security and privacy issues of edns-client-subnet", August 2013, <https://00f.net/2013/08/07/edns-client-subnet/>.

[ditl] CAIDA, "A Day in the Life of the Internet (DITL)", 2002, <http://www.caida.org/projects/ditl/>.

[dns-footprint] Stoner, E., "DNS Footprint of Malware", OARC Workshop, October 2010, <https://www.dns-oarc.net/files/workshop-201010/OARC-ers-20101012.pdf>.

[dnschanger]
Wikipedia, "DNSChanger", October 2013, <https://en.wikipedia.org/w/index.php?title=DNSChanger&oldid=578749672>.

[dnsmezzo]
Bortzmeyer, S., "DNSmezzo", 2009, <http://www.dnsmezzo.net/>.

[fangming-hori-sakurai]
Fangming, Z., Hori, Y., and K. Sakurai, "Analysis of Privacy Disclosure in DNS Query", 2007 International Conference on Multimedia and Ubiquitous Engineering (MUE 2007), Seoul, Korea, ISBN: 0-7695-2777-9, pp. 952-957, DOI 10.1109/MUE.2007.84, April 2007, <http://dl.acm.org/citation.cfm?id=1262690.1262986>.

[federrath-fuchs-herrmann-piosecny]
Federrath, H., Fuchs, K., Herrmann, D., and C. Piosecny, "Privacy-Preserving DNS: Analysis of Broadcast, Range Queries and Mix-based Protection Methods", Computer Security ESORICS 2011, Springer, page(s) 665-683, ISBN 978-3-642-23821-5, 2011, <https://svs.informatik.uni-hamburg.de/publications/2011/2011-09-14_FFHP_PrivacyPreservingDNS_ESORICS2011.pdf>.

[firefox] Deckelmann, , "What's next in making Encrypted DNS-over-HTTPS the Default", September 2019, <https://blog.mozilla.org/futurereleases/2019/09/06/whats-next-in-making-dns-over-https-the-default/>.

[grangeia.snooping]
Grangeia, L., "DNS Cache Snooping or Snooping the Cache for Fun and Profit", 2005, <https://www.semanticscholar.org/paper/Cache-Snooping-or-Snooping-the-Cache-for-Fun-and-1-Grangeia/9b22f606e10b3609eafbdcbfc9090b63be8778c3>.

[herrmann-reidentification]
Herrmann, D., Gerber, C., Banse, C., and H. Federrath, "Analyzing Characteristic Host Access Patterns for Re-Identification of Web User Sessions", DOI 10.1007/978-3-642-27937-9_10, 2012, <http://epub.uni-regensburg.de/21103/1/Paper_PUL_nordsec_published.pdf>.

[I-D.ietf-dnsop-resolver-information]
Sood, P., Arends, R., and P. Hoffman, "DNS Resolver Information Self-publication", draft-ietf-dnsop-resolver-information-00 (work in progress), August 2019.

[I-D.ietf-dprive-bcp-op]
Dickinson, S., Overeinder, B., Rijswijk-Deij, R., and A. Mankin, "Recommendations for DNS Privacy Service Operators", draft-ietf-dprive-bcp-op-03 (work in progress), July 2019.

[I-D.ietf-quic-transport]
Iyengar, J. and M. Thomson, "QUIC: A UDP-Based Multiplexed and Secure Transport", draft-ietf-quic-transport-23 (work in progress), September 2019.

[I-D.ietf-tls-sni-encryption]
Huitema, C. and E. Rescorla, "Issues and Requirements for SNI Encryption in TLS", draft-ietf-tls-sni-encryption-06 (work in progress), September 2019.

[I-D.livingood-doh-implementation-risks-issues]
Livingood, J., Antonakakis, M., Sleigh, B., and A. Winfield, "Centralized DNS over HTTPS (DoH) Implementation Issues and Risks", draft-livingood-doh-implementation-risks-issues-04 (work in progress), September 2019.

[morecowbell]
Grothoff, C., Wachs, M., Ermert, M., and J. Appelbaum, "NSA's MORECOWBELL: Knell for DNS", GNUnet e.V., January 2015, <https://pdfs.semanticscholar.org/2610/2b99bdd6a258a98740af8217ba8da8a1e4fa.pdf>.

[packetq] DNS-OARC, "PacketQ, a simple tool to make SQL-queries against PCAP-files", 2011, <https://github.com/DNS-OARC/PacketQ>.

[passive-dns]
Weimer, F., "Passive DNS Replication", April 2005, <https://www.first.org/conference/2005/papers/florian-weimer-slides-1.pdf>.

[pitfalls-of-dns-encrption]
Shulman, H., "Pretty Bad Privacy:Pitfalls of DNS Encryption", <https://dl.acm.org/citation.cfm?id=2665959>.

[prism] Wikipedia, "PRISM (surveillance program)", July 2015, <https://en.wikipedia.org/w/index.php?title=PRISM_(surveillance_program)&oldid=673789455>.

[RFC4033] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "DNS Security Introduction and Requirements", RFC 4033, DOI 10.17487/RFC4033, March 2005, <https://www.rfc-editor.org/info/rfc4033>.

[RFC5155] Laurie, B., Sisson, G., Arends, R., and D. Blacka, "DNS Security (DNSSEC) Hashed Authenticated Denial of Existence", RFC 5155, DOI 10.17487/RFC5155, March 2008, <https://www.rfc-editor.org/info/rfc5155>.

[RFC5936] Lewis, E. and A. Hoenes, Ed., "DNS Zone Transfer Protocol (AXFR)", RFC 5936, DOI 10.17487/RFC5936, June 2010, <https://www.rfc-editor.org/info/rfc5936>.

[RFC6269] Ford, M., Ed., Boucadair, M., Durand, A., Levis, P., and P. Roberts, "Issues with IP Address Sharing", RFC 6269, DOI 10.17487/RFC6269, June 2011, <https://www.rfc-editor.org/info/rfc6269>.

[RFC7413] Cheng, Y., Chu, J., Radhakrishnan, S., and A. Jain, "TCP Fast Open", RFC 7413, DOI 10.17487/RFC7413, December 2014, <https://www.rfc-editor.org/info/rfc7413>.

[RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <https://www.rfc-editor.org/info/rfc7525>.

[RFC7624] Barnes, R., Schneier, B., Jennings, C., Hardie, T., Trammell, B., Huitema, C., and D. Borkmann, "Confidentiality in the Face of Pervasive Surveillance: A Threat Model and Problem Statement", RFC 7624, DOI 10.17487/RFC7624, August 2015, <https://www.rfc-editor.org/info/rfc7624>.

[RFC7858] Hu, Z., Zhu, L., Heidemann, J., Mankin, A., Wessels, D., and P. Hoffman, "Specification for DNS over Transport Layer Security (TLS)", RFC 7858, DOI 10.17487/RFC7858, May 2016, <https://www.rfc-editor.org/info/rfc7858>.

[RFC7871] Contavalli, C., van der Gaast, W., Lawrence, D., and W. Kumari, "Client Subnet in DNS Queries", RFC 7871, DOI 10.17487/RFC7871, May 2016, <https://www.rfc-editor.org/info/rfc7871>.

[RFC7873] Eastlake 3rd, D. and M. Andrews, "Domain Name System (DNS) Cookies", RFC 7873, DOI 10.17487/RFC7873, May 2016, <https://www.rfc-editor.org/info/rfc7873>.

[RFC7929] Wouters, P., "DNS-Based Authentication of Named Entities (DANE) Bindings for OpenPGP", RFC 7929, DOI 10.17487/RFC7929, August 2016, <https://www.rfc-editor.org/info/rfc7929>.

[RFC8446] Rescorla, E., "The Transport Layer Security (TLS) Protocol Version 1.3", RFC 8446, DOI 10.17487/RFC8446, August 2018, <https://www.rfc-editor.org/info/rfc8446>.

[RFC8484] Hoffman, P. and P. McManus, "DNS Queries over HTTPS (DoH)", RFC 8484, DOI 10.17487/RFC8484, October 2018, <https://www.rfc-editor.org/info/rfc8484>.

[RFC8499] Hoffman, P., Sullivan, A., and K. Fujiwara, "DNS Terminology", BCP 219, RFC 8499, DOI 10.17487/RFC8499, January 2019, <https://www.rfc-editor.org/info/rfc8499>.

[ripe-atlas-turkey] Aben, E., "A RIPE Atlas View of Internet Meddling in Turkey", March 2014, <https://labs.ripe.net/Members/emileaben/a-ripe-atlas-view-of-internet-meddling-in-turkey>.

[ripe-qname-measurements] University of Twente, "Making the DNS More Private with QNAME Minimisation", April 2019, <https://labs.ripe.net/Members/wouter_de_vries/make-dns-a-bit-more-private-with-qname-minimisation>.

[sidn-entrada] Hesselman, C., Jansen, J., Wullink, M., Vink, K., and M. Simon, "A privacy framework for 'DNS big data' applications", November 2014, <https://www.sidnlabs.nl/downloads/yBW6hBoaSZe4m6GJc_0b7w/2211058ab6330c7f3788141ea19d3db7/SIDN_Labs_Privacyraamwerk_Position_Paper_V1.4_ENG.pdf>.

[thomas-ditl-tcp]
Thomas, M. and D. Wessels, "An Analysis of TCP Traffic in Root Server DITL Data", DNS-OARC 2014 Fall Workshop, October 2014, <https://indico.dns-oarc.net/event/20/session/2/contribution/15/material/slides/1.pdf>.

[tor-leak]
Tor, "DNS leaks in Tor", 2013, <https://www.torproject.org/docs/faq.html.en#WarningsAboutSOCKSandDNSInformationLeaks>.

[yanbin-tsudik]
Yanbin, L. and G. Tsudik, "Towards Plugging Privacy Leaks in the Domain Name System", October 2009, <http://arxiv.org/abs/0910.2472>.

9.3. URIs

[1] https://lists.dns-oarc.net/pipermail/dns-operations/2016-January/014141.html

[2] http://netres.ec/?b=11B99BD

[3] https://www.researchgate.net/publication/320322146_DNS-DNS_DNS-based_De-NAT_Scheme

[4] https://developers.google.com/speed/public-dns/privacy

[5] https://mailarchive.ietf.org/arch/browse/static/add

[6] https://www.encrypted-dns.org

[7] https://www.eugdpr.org/the-regulation.html

Authors' Addresses

Stephane Bortzmeyer
AFNIC
1, rue Stephenson
Montigny-le-Bretonneux
France 78180

Email: bortzmeyer+ietf@nic.fr

Sara Dickinson
Sinodun IT
Magdalen Centre
Oxford Science Park
Oxford OX4 4GA
United Kingdom

Email: sara@sinodun.com