Internet Engineering Task Force (IETF)
Request for Comments: 9076
Obsoletes: 7626
Category: Informational
ISSN: 2070-1721

T. Wicinski, Ed.

July 2021

DNS Privacy Considerations

Abstract

This document describes the privacy issues associated with the use of the DNS by Internet users. It provides general observations about typical current privacy practices. It is intended to be an analysis of the present situation and does not prescribe solutions. This document obsoletes RFC 7626.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are candidates for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc9076.

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 is 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.

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, 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 servers will then return the answers. 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.

DNS relies heavily on caching, so the algorithm described above is actually a bit more complicated, and not all questions are sent to the authoritative name servers. If the stub resolver asks the recursive resolver a few seconds later, "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, almost all this DNS traffic is currently sent 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. Some recent analysis on the service quality of encrypted DNS traffic can be found in [dns-over-encryption].

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, although new solutions are still emerging [RFC9000] [DPRIVE-DNSOQUIC].

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 for different reasons. 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 a 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 service 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 send additional requests to turn the name servers' names into IP addresses. Similarly, even if glue records are returned, a careful recursive server will send tertiary requests to verify the IP addresses of those records.

It can also be noted 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 significant 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.

Privacy-related terminology is from [RFC6973]. This document obsoletes [RFC7626].

2. Scope

This document focuses mostly on the study of privacy risks for the end user (the one performing DNS requests). The risks of pervasive surveillance [RFC7258] are considered as well as risks coming from a more focused surveillance. In this document, the term "end user" is used as defined in [RFC8890].

This document does not attempt a comparison of specific privacy protections provided by individual networks or organizations; it makes only general observations about typical current practices.

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

Privacy risks for recursive operators (including access providers and operators in enterprise networks) such as leakage of private namespaces or blocklists are out of scope for this document.

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

The privacy risks associated with the use of other protocols that make use of DNS information are not considered here.

3. Risks

The following four sections outline the privacy considerations associated with different aspects of the DNS for the end user. When reading these sections, it needs to be kept in mind that many of the considerations (for example, recursive resolver and transport protocol) can be specific to the network context that a device is using at a given point in time. A user may have many devices, and each device might utilize many different networks (e.g., home, work, public, or cellular) over a period of time or even concurrently. An exhaustive analysis of the privacy considerations for an individual user would need to take into account the set of devices used and the multiple dynamic contexts of each device. This document does not attempt such a complex analysis; instead, it presents an overview of the various considerations that could form the basis of such an analysis.

4. Risks in the DNS Data

4.1. The Public Nature of DNS Data

It has been stated that "the data in the DNS is public". This sentence makes sense for an Internet-wide lookup system, and 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 nonexistence). In other words: one needs to know what to ask for in order to receive a response. There are many ways in which supposedly "private" resources currently leak. A few examples are DNSSEC NSEC zone walking [RFC4470], passive DNS services [passive-dns], etc. The zone transfer QTYPE [RFC5936] is often blocked or restricted to authenticated/authorized access to enforce this difference (and maybe for other reasons).

Another difference between the DNS data and a particular DNS 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; those transactions are not / should not be public. A single transaction reveals both the originator of the query and the query contents; this potentially leaks sensitive information about a specific user. A typical example from outside the DNS world is that the website of Alcoholics Anonymous is public but the fact that you visit it should not be. Furthermore, the ability to link queries reveals information about individual use patterns.

4.2. 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), 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 they probably want to send 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 websites 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 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 exchanging emails 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 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 typically the IP address of the recursive resolver that, in a way, "hides" the real user. However, hiding does not always work. The edns-client-subnet (ECS) EDNS0 option [RFC7871] is sometimes used (see one privacy analysis in [denis-edns-client-subnet]). Sometimes the end user has a personal recursive resolver on their machine. In both cases, the IP address originating queries to the authoritative server 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 general, although [RFC7721] does discuss privacy considerations for IPv6 address generation mechanisms. 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 an address-sharing scheme.) However, for both IPv4 and IPv6 addresses, it is important to note that source addresses are propagated with queries via the ECS option and comprise metadata about the host, user, or application that originated them.

4.2.1. Data in the DNS Payload

At the time of writing, there are no standardized client identifiers contained in the DNS payload itself (ECS, as described in [RFC7871], is widely used; however, [RFC7871] is only an Informational RFC).

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), but they are not designed to actively track users (like HTTP cookies).

There are anecdotal accounts of Media Access Control (MAC) addresses (https://lists.dns-oarc.net/pipermail/dns-operations/2016-January/014143.html) and even user names being inserted in nonstandard EDNS(0) options [RFC6891] for stub-to-resolver communications to support proprietary functionality implemented at the resolver (e.g., parental filtering).

4.3. 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 countermeasures have already been developed and deployed [cache-snooping-defence].

5. Risks on the Wire

5.1. 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, the HTTP traffic will be encrypted, but the DNS exchange prior to it will not be. When other protocols become more and more privacy aware and secured against surveillance (e.g., [RFC8446], [RFC9000]), 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 ongoing work attempting to encrypt the Server Name Identification (SNI) in the TLS handshake [RFC8744], which is one of the last remaining non-DNS cleartext identifiers of a connection target.

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 device is configured. By order of increasing attack surface:

* The recursive resolver can be on the end user's device. 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 will expose data to authoritative resolvers as discussed in Section 6.2.

* The recursive resolver may be at the local network edge. For many/most enterprise networks and for some residential networks, 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 network. For most residential networks and potentially other networks, the typical case is for the user's device to be configured (typically automatically through DHCP or relay agent options) with the addresses of the DNS proxy in the Customer Premises Equipment (CPE), which in turn points to 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 (or a privately run DNS resolver hosted on the public Internet). Some machines may be configured to use public DNS resolvers such as those operated by Google Public DNS or OpenDNS. The 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 user's connection and the public DNS service. It can be noted that if the user selects a single resolver with a small client population (even when using an encrypted transport), it can actually serve to aid tracking of that user as they move across network environments.

It is also noted that, typically, a device connected _only_ to a modern cellular network is

* directly configured with only the recursive resolvers of the IAP and

* afforded some level of protection against some types of eavesdropping for all traffic (including DNS traffic) due to the cellular network link-layer encryption.

The attack surface for this specific scenario is not considered here.

5.2. Encrypted Transports

The use of encrypted transports directly mitigates passive surveillance of the DNS payload; however, some privacy attacks are still 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 cleartext/observable features. These issues are not specific to DNS, but DNS traffic is susceptible to these attacks when using specific transports.

Some general examples exist; for example, certain studies highlight that the OS fingerprint values (http://netres.ec/?b=11B99BD) of IPv4 TTL, IPv6 Hop Limit, or TCP Window size can be used to fingerprint client OSes or that various techniques can be used to de-NAT DNS queries [dns-de-nat].

Note that even when using encrypted transports, the use of cleartext transport options to decrease latency can provide correlation of a user's connections, e.g., using TCP Fast Open [RFC7413].

Implementations that support encrypted transports also commonly reuse connections for multiple DNS queries to optimize performance (e.g., via DNS pipelining or HTTPS multiplexing). Default configuration options for encrypted transports could, in principle, fingerprint a specific client application. For example:

* TLS version or cipher suite selection

* session resumption

* the maximum number of messages to send and

* a maximum connection time before closing a connections and reopening.

If libraries or applications offer user configuration of such options (e.g., [getdns]), then they could, in principle, help to identify a specific user. Users may want to use only the defaults to avoid this issue.

While there are known attacks on older versions of TLS, the most recent recommendations [RFC7525] and the development of TLS 1.3 [RFC8446] largely mitigate those.

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

6. Risks 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 [RFC6973] again, "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]).

6.1. 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.

6.1.1. Resolver 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-provided local network recursive resolver. The choice by a user to join a particular network (e.g., by physically plugging in a cable or selecting a network in an OS dialogue) typically updates a number of system resources -- these can include IP addresses, the availability of IPv4/IPv6, DHCP server, and DNS resolver. These individual changes, including the change in DNS

resolver, are not normally communicated directly to the user by the OS when the network is joined. The choice of network has historically determined the default system DNS resolver selection; the two are directly coupled in this model.

The vast majority of users do not change their default system DNS settings and so implicitly accept the network settings for the DNS. The network resolvers have therefore historically been the sole destination for all of the DNS queries from a device. These resolvers may have varied 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 the network.

All major OSes expose the system DNS settings and allow users to manually override them if desired.

More recently, some networks and users have actively chosen to use a large public resolver, e.g., Google Public DNS (https://developers.google.com/speed/public-dns), Cloudflare (https://developers.cloudflare.com/1.1.1.1/setting-up-1.1.1.1/), or Quad9 (https://www.quad9.net). 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 user can get an idea of the data collected by such operators by reading one, e.g., Google Public DNS - Your Privacy (https://developers.google.com/speed/public-dns/privacy).

In general, as with many other protocols, issues around centralization also arise with DNS. The picture is fluid with several competing factors contributing, where these factors can also vary by geographic region. These include:

* ISP outsourcing, including to third-party and public resolvers

* regional market domination by one or only a few ISPs

* applications directing DNS traffic by default to a limited subset of resolvers (see Section 6.1.1.2)

An increased proportion of the global DNS resolution traffic being served by only a few entities means that the privacy considerations for users are highly dependent on the privacy policies and practices of those entities. Many of the issues around centralization are discussed in [centralisation-and-data-sovereignty].

6.1.1.1. Dynamic Discovery of DoH and Strict DoT

While support for opportunistic DoT can be determined by probing a resolver on port 853, there is currently no standardized discovery mechanism for DoH and Strict DoT servers.

This means that clients that might want to dynamically discover such encrypted services, and where users are willing to trust such services, are not able to do so. At the time of writing, efforts to

provide standardized signaling mechanisms to discover the services offered by local resolvers are in progress [DNSOP-RESOLVER]. Note that an increasing number of ISPs are deploying encrypted DNS; for example, see the Encrypted DNS Deployment Initiative [EDDI].

6.1.1.2. Application-Specific Resolver Selection

An increasing number of applications are offering application-specific encrypted DNS resolution settings, rather than defaulting to using only the system resolver. A variety of heuristics and resolvers are available in different applications, including hard-coded lists of recognized DoH/DoT servers.

Generally, users are not aware of application-specific DNS settings and may not have control over those settings. To address these limitations, users will only be aware of and have the ability to control such settings if applications provide the following functions:

* communicate the change clearly to users when the default application resolver changes away from the system resolver

* provide configuration options to change the default application resolver, including a choice to always use the system resolver

* provide mechanisms for users to locally inspect, selectively forward, and filter queries (either via the application itself or use of the system resolver)

Application-specific changes to default destinations for users' DNS queries might increase or decrease user privacy; it is highly dependent on the network context and the application-specific default. This is an area of active debate, and the IETF is working on a number of issues related to application-specific DNS settings.

6.1.2. Active Attacks on Resolver Configuration

The previous 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, there can be active attackers in the network.

The Internet Threat model, as described in [RFC3552], assumes that the attacker controls the network. Such an attacker can completely control any insecure DNS resolution, both passively monitoring the queries and responses and substituting their own responses. Even if encrypted DNS such as DoH or DoT is used, unless the client has been configured in a secure way with the server identity, an active attacker can impersonate the server. This implies that opportunistic modes of DoH/DoT as well as modes where the client learns of the DoH/DoT server via in-network mechanisms such as DHCP are vulnerable to attack. In addition, if the client is compromised, the attacker can replace the DNS configuration with one of its own choosing.

6.1.3. Blocking of DNS Resolution Services

User privacy can also be at risk if there is blocking of access to remote recursive servers that offer encrypted transports, e.g., 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 blocking of specific IP addresses could restrict the resolvers available to the user. The extent of the risk to user privacy is highly 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.

As a matter of policy, some recursive resolvers use their position in the query path to selectively block access to certain DNS records. This is a form of rendezvous-based blocking as described in Section 4.3 of [RFC7754]. Such blocklists often include servers known to be used for malware, bots, or other security risks. In order to prevent circumvention of their blocking policies, some networks also block access to resolvers with incompatible policies.

It is also noted that attacks on remote resolver services, e.g., DDoS, could force users to switch to other services that do not offer encrypted transports for DNS.

6.1.4. Encrypted Transports and Recursive Resolvers

6.1.4.1. 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 described in Section 5.2, use of session-based encrypted transports (TCP/TLS) can expose correlation data about users.

6.1.4.2. DoH-Specific Considerations

DoH inherits the full privacy properties of the HTTPS stack and as a consequence introduces new privacy considerations when compared with DNS over UDP, TCP, or TLS [RFC7858]. Section 8.2 of [RFC8484] describes the privacy considerations in the server of the DoH protocol.

A brief summary of some of the issues includes the following:

* HTTPS presents new considerations for correlation, such as explicit HTTP cookies and implicit fingerprinting of the unique set and ordering of HTTP request header fields.

* The User-Agent and Accept-Language request header fields often convey specific information about the client version or locale.

* Utilizing the full set of HTTP features enables DoH to be more than an HTTP tunnel, but it is at the cost of opening up implementations to the full set of privacy considerations of HTTP.

* Implementations are advised to expose the minimal set of data

needed to achieve the desired feature set.

[RFC8484] specifically makes selection of HTTPS functionality vs. privacy an implementation choice. 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 HTTP headers, and there might be others that provide feature-rich data flows where the low-level origin of the DNS query is easily identifiable. Some implementations have, in fact, chosen to restrict the use of the User-Agent header so that resolver operators cannot identify the specific application that is originating the DNS queries.

Privacy-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.

6.2. In the Authoritative Name Servers

Unlike what happens for recursive resolvers, the 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 minimization [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 user often has some legal/contractual link with the recursive resolver (they have chosen the IAP, or they have 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 or may not be honest; in any case, it will have to follow its local laws. For example, requests to a given ccTLD may go to servers managed by organizations outside of the ccTLD's country. 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 (https://www.alexa.com/topsites), one DNS provider hosts 10% of the domains today. The ten most important DNS providers together host one-third of all domains. With the control (or the ability to sniff the traffic) of a few name servers, you can gather a lot of information.

7. Other Risks

7.1. Re-identification and Other Inferences

An observer has access not only to the data they directly collect but also to the results of various inferences about this data. The term "observer" here is used very generally; for example, the observer might passively observe cleartext DNS traffic or be in the network that is actively attacking the user by redirecting 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 has a document [RFC7624] that considers such inference-based attacks in a more general framework.

7.2. More Information

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

8. 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 greater 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 services [passive-dns] allow reconstruction of the data of sometimes an entire zone. Well-known passive DNS services keep only the DNS responses and not the source IP address of the client, precisely for privacy reasons. Other passive DNS services may not be so careful. And there are still 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 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 (https://theintercept.com/document/2014/03/12/nsa-gchqs-quantumtheory-hacking-tactics/) 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.

9. Legalities

To our knowledge, there are no specific privacy laws for DNS data in any country. Interpreting general privacy laws, like the European Union's [data-protection-directive] or GDPR (https://gdpr.eu/tag/gdpr/), in the context of DNS traffic data is not an easy task, and there is no known court precedent. See an interesting analysis in [sidn-entrada].

10. 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" [RFC8932].

11. IANA Considerations

This document has no IANA actions.

12. References

12.1. 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>.

12.2. Informative References

[aeris-dns] Vinot, N., "Vie privée: et le DNS alors? [Privacy: what about DNS?]", February 2015, <https://blog.imirhil.fr/vie-privee-et-le-dns-alors.html>.

[cache-snooping-defence] ISC, "DNS Cache snooping - should I be concerned?", October 2018, <https://kb.isc.org/docs/aa-00482>.

[castillo-garcia] Castillo-Perez, S. and J. Garcia-Alfaro, "Anonymous Resolution of DNS Queries", Lecture Notes in Computer Science, Vol. 5332, DOI 10.1007/978-3-540-88873-4_5, 2008, <https://dl.acm.org/doi/10.1007/978-3-540-88873-4_5>.

[centralisation-and-data-sovereignty] De Filippi, P. and S. McCarthy, "Cloud Computing: Centralization and Data Sovereignty", European Journal of Law and Technology, Vol. 3, No. 2, October 2012, <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2167372>.

[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", May 2013, <https://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 Parliament and of the Council of 24 October 1995 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. 31-50, November 1995, <https://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, No. 5, DOI 10.1145/1452335.1452341, October 2008, <https://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)", <https://www.caida.org/projects/ditl/>.

[dns-de-nat]
Orevi, L., Herzberg, A., Zlatokrilov, H., and D. Sigron, "DNS-DNS: DNS-based De-NAT Scheme", January 2017, <https://www.researchgate.net/publication/320322146_DNS-DNS_DNS-based_De-NAT_Scheme>.

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

[dns-over-encryption]
Lu, C., Liu, B., Li, Z., Hao, S., Duan, H., Zhang, M., Leng, C., Liu, Y., Zhang, Z., and J. Wu, "An End-to-End, Large-Scale Measurement of DNS-over-Encryption: How Far Have We Come?", IMC '19: Proceedings of the Internet Measurement Conference, pp. 22-35, DOI 10.1145/3355369.3355580, October 2019, <https://dl.acm.org/citation.cfm?id=3355369.3355580>.

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

[DNSOP-RESOLVER]
Sood, P., Arends, R., and P. Hoffman, "DNS Resolver Information Self-publication", Work in Progress, Internet-Draft, draft-ietf-dnsop-resolver-information-01, 11 February 2020, <https://datatracker.ietf.org/doc/html/draft-ietf-dnsop-resolver-information-01>.

[DPRIVE-DNSOQUIC]
Huitema, C., Dickinson, S., and A. Mankin, "Specification of DNS over Dedicated QUIC Connections", Work in Progress, Internet-Draft, draft-ietf-dprive-dnsoquic-03, 12 July 2021, <https://datatracker.ietf.org/doc/html/draft-ietf-dprive-dnsoquic-03>.

[EDDI] EDDI, "Encrypted DNS Deployment Initiative", <https://www.encrypted-dns.org>.

[fangming-hori-sakurai]
Zhao, F., Hori, Y., and K. Sakurai, "Analysis of Privacy Disclosure in DNS Query", MUE '07: Proceedings of the 2007 International Conference on Multimedia and Ubiquitous Engineering, pp. 952-957, DOI 10.1109/MUE.2007.84, ISBN 0-7695-2777-9, April 2007, <https://dl.acm.org/citation.cfm?id=1262690.1262986>.

[federrath-fuchs-herrmann-piosecny]
Federrath, H., Fuchs, K.-P., Herrmann, D., and C. Piosecny, "Privacy-Preserving DNS: Analysis of Broadcast, Range Queries and Mix-based Protection Methods", ESORICS 2011, pp. 665-683, DOI 10.1007/978-3-642-23822-2_36, ISBN 978-3-642-23822-2, 2011, <https://svs.informatik.uni-hamburg.de/publications/2011/2011-09-14_FFHP_PrivacyPreservingDNS_ESORICS2011.pdf>.

[getdns] "getdns", <https://getdnsapi.net>.

[grangeia.snooping]
Grangeia, L., "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", Lecture Notes in Computer Science, Vol. 7127, DOI 10.1007/978-3-642-27937-9_10, 2012, <https://epub.uni-regensburg.de/21103/1/Paper_PUL_nordsec_published.pdf>.

[morecowbell]
Grothoff, C., Wachs, M., Ermert, M., and J. Appelbaum,

"NSA's MORECOWBELL: Knell for DNS", January 2015, <https://pdfs.semanticscholar.org/2610/2b99bdd6a258a98740af8217ba8da8a1e4fa.pdf>.

[packetq] DNS-OARC, "A tool that provides a basic SQL-frontend to PCAP-files", Release 1.4.3, commit 29a8288, October 2020, <https://github.com/DNS-OARC/PacketQ>.

[passive-dns] Weimer, F., "Passive DNS Replication", 17th Annual FIRST Conference, April 2005, <https://www.first.org/conference/2005/papers/florian-weimer-slides-1.pdf>.

[pitfalls-of-dns-encryption] Shulman, H., "Pretty Bad Privacy: Pitfalls of DNS Encryption", WPES '14: Proceedings of the 13th Workshop on Privacy in the Electronic Society, pp. 191-200, DOI 10.1145/2665943.2665959, November 2014, <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>.

[RFC3552] Rescorla, E. and B. Korver, "Guidelines for Writing RFC Text on Security Considerations", BCP 72, RFC 3552, DOI 10.17487/RFC3552, July 2003, <https://www.rfc-editor.org/info/rfc3552>.

[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>.

[RFC4470] Weiler, S. and J. Ihren, "Minimally Covering NSEC Records and DNSSEC On-line Signing", RFC 4470, DOI 10.17487/RFC4470, April 2006, <https://www.rfc-editor.org/info/rfc4470>.

[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>.

[RFC6891] Damas, J., Graff, M., and P. Vixie, "Extension Mechanisms for DNS (EDNS(0))", STD 75, RFC 6891,

DOI 10.17487/RFC6891, April 2013,
<https://www.rfc-editor.org/info/rfc6891>.

[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>.

[RFC7626] Bortzmeyer, S., "DNS Privacy Considerations", RFC 7626, DOI 10.17487/RFC7626, August 2015, <https://www.rfc-editor.org/info/rfc7626>.

[RFC7721] Cooper, A., Gont, F., and D. Thaler, "Security and Privacy Considerations for IPv6 Address Generation Mechanisms", RFC 7721, DOI 10.17487/RFC7721, March 2016, <https://www.rfc-editor.org/info/rfc7721>.

[RFC7754] Barnes, R., Cooper, A., Kolkman, O., Thaler, D., and E. Nordmark, "Technical Considerations for Internet Service Blocking and Filtering", RFC 7754, DOI 10.17487/RFC7754, March 2016, <https://www.rfc-editor.org/info/rfc7754>.

[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>.

[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>.

[RFC8744] Huitema, C., "Issues and Requirements for Server Name Identification (SNI) Encryption in TLS", RFC 8744, DOI 10.17487/RFC8744, July 2020, <https://www.rfc-editor.org/info/rfc8744>.

[RFC8890] Nottingham, M., "The Internet is for End Users", RFC 8890, DOI 10.17487/RFC8890, August 2020, <https://www.rfc-editor.org/info/rfc8890>.

[RFC8932] Dickinson, S., Overeinder, B., van Rijswijk-Deij, R., and A. Mankin, "Recommendations for DNS Privacy Service Operators", BCP 232, RFC 8932, DOI 10.17487/RFC8932, October 2020, <https://www.rfc-editor.org/info/rfc8932>.

[RFC9000] Iyengar, J., Ed. and M. Thomson, Ed., "QUIC: A UDP-Based Multiplexed and Secure Transport", RFC 9000, DOI 10.17487/RFC9000, May 2021, <https://www.rfc-editor.org/info/rfc9000>.

[ripe-qname-measurements] de Vries, W., "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, "Tor FAQs: I keep seeing these warnings about SOCKS and DNS information leaks. Should I worry?", <https://www.torproject.org/docs/

faq.html.en#WarningsAboutSOCKSandDNSInformationLeaks>.

[yanbin-tsudik]
Yanbin, L. and G. Tsudik, "Towards Plugging Privacy Leaks in Domain Name System", June 2010, <https://arxiv.org/abs/0910.2472>.

Appendix A. Updates since RFC 7626

Many references were updated. Discussions of encrypted transports, including DoT and DoH, and sections on DNS payload, authentication of servers, and blocking of services were added. With the publishing of [RFC7816] on QNAME minimization, text, references, and initial attempts to measure deployment were added to reflect this. The text and references on the Snowden revelations were updated.

The "Risks Overview" section was changed to "Scope" to help clarify the risks being considered. Text on cellular network DNS, blocking, and security was added. Considerations for recursive resolvers were collected and placed together. A discussion on resolver selection was added.

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, 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. Thanks to Vittorio Bertola and Mohamed Boucadair for a detailed review of the -bis. And thanks to the IESG members for the last remarks.

Contributions

Sara Dickinson and Stephane Bortzmeyer were the original authors of the document, and their contribution to the initial draft of this document is greatly appreciated.

Author's Address

Tim Wicinski (editor)
Elkins, WV 26241
United States of America

Email: tjw.ietf@gmail.com