Network Working Group
Internet-Draft
Intended status: Informational
Expires: October 29, 2014

S. Bortzmeyer
AFNIC

April 27, 2014

DNS privacy considerations
draft-bortzmeyer-dnsop-dns-privacy-02

Abstract

This document describes the privacy issues associated with the use of the DNS by Internet users. It is intended to be mostly an analysis of the present situation, in the spirit of section 8 of [RFC6973] and it does not prescribe solutions.

Discussions of the document should take place on the dns-privacy mailing list [dns-privacy].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on October 29, 2014.

Table of Contents

1. Introduction

The Domain Name System is specified in [RFC1034] and [RFC1035]. It is one of the most important infrastructure components of the Internet and one of the most often ignored or misunderstood. Almost every activity on the Internet starts with a DNS query (and often several). Its use has many privacy implications and we try to give here a comprehensive and accurate list.

Let us start with a small reminder of the way the DNS works (with some simplifications). A client, the stub resolver, issues a DNS query to a server, the resolver (also called caching resolver or full resolver or recursive name server). For instance, the query is "What are the AAAA records for www.example.com?". AAAA is the qtype (Query Type) and www.example.com the qname (Query Name). To get the answer, the resolver will query first the root nameservers, which will, most of the times, send a referral. Here, the referral will be to .com nameservers. In turn, they will send a referral to the example.com nameservers, which will provide the answer. The root name servers, the name servers of .com and those 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. Unlike what many "DNS for dummies" articles say, 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?". So, the DNS leaks more information than it should.

Because the DNS uses caching heavily, not all questions are sent to the authoritative name servers. If the stub resolver, a few seconds later, asks to the resolver "What are the SRV records of _xmpp-server._tcp.example.com?", the 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 resolver, unlike the authoritative servers, sees everything.

Almost all the DNS queries are today sent over UDP, and this has practical consequences if someone thinks of encrypting this traffic (some encryption solutions are typically done for TCP, not UDP).

I should be noted to that DNS resolvers sometimes forward requests to bigger machines, with a larger and more shared cache, the forwarders. From the point of view of privacy, forwarders are like resolvers, except that the caching in the resolver before them decreases the amount of data they can see.

Another important point to keep in mind when analyzing the privacy issues of DNS is the mix of many sort of DNS requests received by a server. Let's assume the eavesdropper want to know which Web page is visited by an user. For a typical Web page displayed by the user, there are three sorts of DNS requests:

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

Secondary requests: these are the requests performed by the user agent (here, the Web browser) without any direct involvment or knowledge of the user. For the Web, they are triggered by included content, CSS sheets, JavaScript code, embedded images, etc. In some cases, there can be dozens of domain names in a single page.

Tertiary requests: these are the 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 is not returned, the resolver will have to do tertiary requests to turn name servers' named into IP addresses.

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

2. Risks

This draft focuses mostly on the study of privacy risks for the end-user (the one performing DNS requests). Privacy risks for the holder of a zone (the risk that someone gets the data) are discussed in [RFC5936]. Non-privacy risks (such as cache poisoning) are out of scope.

2.1. 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 data and meta data that deserve a more detailed look. First, access control lists and private name spaces nonwithstanding, 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 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, more dubious reasons).

Another differentiation to be applied is between the DNS data as mentioned above and a particular transaction, most prominently but not limited to a DNS name lookup. The fact that the results of a DNS query are public within the boundaries described in the previous paragraph and therefore might have no confidentiality requirements does not imply the same for a single or a sequence of transactions. A typical example from outside the DNS world: the Web site of Alcoholics Anonymous is public, the fact that you visit it should not be.

2.2. Data in the DNS request

The DNS request includes many fields but two of them seem specially 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 + may be source port", because the port is also in the request and can be used to sort out several users sharing an IP address (CGN for instance).

The qname is the full name sent by the original user. It gives information about what the user does ("What are the MX records of example.net?" means he probably wants to send email to someone at example.net, which may be a domain used by only a few persons and

therefore very revealing). Some qnames are more sensitive than others. For instance, querying the A record of google-analytics.com reveals very little (everybody visits Web sites which use Google Analytics) but querying the A record of www.verybad.example where verybad.example is the domain of an illegal or very offensive organization may create more problems for the user. Another example is when the qname embeds the software one uses. For instance, _ldap._tcp.Default-First-Site-Name._sites.gc._msdcs.example.org. Or some BitTorrent clients that query a 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 interesting data in the DNS. At the moment your DNS traffic might reveal that you are doing email but not who with. If your MUA starts looking up PGP keys in the DNS [I-D.wouters-dane-openpgp] then privacy becomes a lot more important. And email is just an example, there will be other really interesting uses for a more secure (in the sense of privacy) DNS.

For the communication between the stub resolver and the resolver, the source IP address is the one of the user's machine. Therefore, all the issues and warnings about collection of IP addresses apply here. For the communication between the 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 a HTTP connection. It is now the IP address of the resolver which, in a way "hides" the real user. However, it does not always work. Sometimes [I-D.vandergaast-edns-client-subnet] is used. Sometimes the end user has a personal resolver on her machine. In that case, the IP address is as sensitive as it is for HTTP.

A note about IP addresses: there is currently no IETF document which describes in detail the privacy issues of IP addressing. In the mean time, 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.) 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.

2.3. Cache snooping

The content of resolvers can reveal data about the clients using it. This information can sometimes be examined by sending DNS queries with RD=0 to inspect cache content, particularly looking at the DNS TTLs. Since this also is a reconnaissance technique for subsequent cache poisoning attacks, some counter measures have already been developed and deployed.

2.4. On the wire

DNS traffic can be seen by an eavesdropper like any other traffic. It is typically not encrypted. (DNSSEC, specified in [RFC4033] explicitely excludes confidentiality from its goals.) So, if an initiator starts a HTTPS communication with a recipient, while the HTTP traffic will be encrypted, the DNS exchange prior to it will not be. When the other protocols will become more or more privacy-aware and secured against surveillance, the DNS risks to become "the weakest link" in privacy.

What also makes the DNS traffic different 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 resolver, or to the authoritative name servers.

The best place, from an eavesdropper's point of view, is clearly between the stub resolvers and the resolvers, because he 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 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 stub resolver to caching resolver connection is limited to that single machine.

The resolver can be in the IAP (Internet Access Provider) 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) with the addresses of the DNS resolver 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 resolvers.

The resolver may also 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 IAP network referenced above.

The resolver can be a public DNS service. Some end users may be configured to use public DNS resolvers such as those operated by Google Public DNS or OpenDNS. The end user may have configured their machine to use these DNS resolvers themselves - or their IAP may choose 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.

2.5. In the servers

Using the terminology of [RFC6973], the DNS servers (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] reference 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 attacker.

Sometimes, these data are kept for a long time and/or distributed to third parties, for research purposes [ditl], for security analysis, or for surveillance tasks. Also, there are observation points in the network which gather DNS data and then make it accessible to third-parties for research or security purposes ("passive DNS [passive-dns]").

2.5.1. In the resolvers

The resolvers see the entire traffic since there is typically no caching before them. They are therefore well situated to observe the traffic. To summarize: your resolver knows a lot about you. The resolver of a large IAP, or a large public resolver can collect data from many users. You may get an idea of the data collected by reading the privacy policy of a big public resolver [1].

2.5.2. In the authoritative name servers

Unlike the resolvers, they are limited by caching. They see only a part of the requests. For aggregated statistics ("what is the percentage of LOC queries?"), it is sufficient but it may prevent an observer to observe everything. Nevertheless, the authoritative name servers sees a part of the traffic and this sample may be sufficient to defeat some privacy expectations.

Also, the end user has typically some legal/contractual link with the resolver (he has chosen the IAP, or he has chosen to use a given public resolver) while he is often not even aware of the role of the authoritative name servers and their observation abilities.

It is an interesting question whether the privacy issues are bigger in the root or in a large TLD. The root sees the traffic for all the TLDs (and the huge amount of traffic for non-existing TLD) but a large TLD has less caching before it.

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 since the authoritative name server will see the IP address of the end client, and not the address of a big resolver shared by many users. This is no longer true if [I-D.vandergaast-edns-client-subnet] is used because, in this case, the authoritative name server sees the original IP prefix or address (depending on the setup).

As of today, all the instances of one root name server, L-root, receive together around 20 000 queries per second. While most of it is junk (errors on the TLD name), it gives an idea of the amount of big data which pours into name servers.

Many domains, including TLD, 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. But it may be surprising for an end-user that requests to a given ccTLD may go to servers managed by organisations outside of the country.

2.5.3. Rogue servers

A rogue DHCP server can direct you to a rogue resolver. Most of the times, it seems to be done to divert traffic, by providing lies for some domain names. But it could be used just to capture the traffic and gather information about you. Same thing for malwares like DNSchanger[dnschanger] which changes the resolver in the machine's configuration.

3. 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) second 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 you're actually interested in.

Many research papers about malware detection use DNS traffic to detect "abnormal" behaviour 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. It is used for many reasons, some good, some bad. It is an example of privacy issue even when no source IP address is kept.

4. Legalities

To our knowledge, there are no specific privacy laws for DNS data. Interpreting general privacy laws like [data-protection-directive] (European Union) in the context of DNS traffic data is not an easy task and it seems there is no court precedent here.

5. Security considerations

This document is entirely about security, more precisely privacy. Possible solutions to the issues described here are discussed in [I-D.bortzmeyer-dnsop-privacy-sol] (qname minimization, local caching resolvers), [I-D.hzhwm-start-tls-for-dns] (encryption of traffic) or in [I-D.wijngaards-dnsop-confidentialdns] (encryption also). Attempts have been made to encrypt the resource record data [I-D.timms-encrypt-naptr].

6. Acknowledgments

Thanks to Nathalie Boulvard and to the CENTR members for the original work which leaded to this draft. Thanks to Ondrej Sury for the interesting discussions. Thanks to Mohsen Souissi for proofreading. Thanks to Dan York, Suzanne Woolf, Tony Finch, Peter Koch and Frank Denis for good written contributions.

7. References

7.1. Normative References

[RFC1034] Mockapetris, P., "Domain names - concepts and facilities", STD 13, RFC 1034, November 1987.

[RFC1035] Mockapetris, P., "Domain names - implementation and specification", STD 13, RFC 1035, November 1987.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC6973] Cooper, A., Tschofenig, H., Aboba, B., Peterson, J., Morris, J., Hansen, M., and R. Smith, "Privacy Considerations for Internet Protocols", RFC 6973, July 2013.

7.2. Informative References

[RFC2181] Elz, R. and R. Bush, "Clarifications to the DNS Specification", RFC 2181, July 1997.

[RFC4033] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "DNS Security Introduction and Requirements", RFC 4033, March 2005.

[RFC5246] Dierks, T. and E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.2", RFC 5246, August 2008.

[RFC5936] Lewis, E. and A. Hoenes, "DNS Zone Transfer Protocol (AXFR)", RFC 5936, June 2010.

[I-D.vandergaast-edns-client-subnet]
Contavalli, C., Gaast, W., Leach, S., and E. Lewis, "Client Subnet in DNS Requests", draft-vandergaast-edns-client-subnet-02 (work in progress), July 2013.

[I-D.bortzmeyer-dnsop-privacy-sol]
Bortzmeyer, S., "Possible solutions to DNS privacy issues", draft-bortzmeyer-dnsop-privacy-sol-00 (work in progress), December 2013.

[I-D.wijngaards-dnsop-confidentialdns]
Wijngaards, W., "Confidential DNS", draft-wijngaards-dnsop-confidentialdns-00 (work in progress), November 2013.

[I-D.timms-encrypt-naptr]
Timms, B., Reid, J., and J. Schlyter, "IANA Registration for Encrypted ENUM", draft-timms-encrypt-naptr-01 (work in progress), July 2008.

[I-D.hzhwm-start-tls-for-dns]
Zi, Z., Zhu, L., Heidemann, J., Mankin, A., and D. Wessels, "Starting TLS over DNS", draft-hzhwm-start-tls-for-dns-00 (work in progress), February 2014.

[I-D.wouters-dane-openpgp]
Wouters, P., "Using DANE to Associate OpenPGP public keys with email addresses", draft-wouters-dane-openpgp-02 (work in progress), February 2014.

[dns-privacy]
IETF, , "The dns-privacy mailing list", March 2014.

[dnsop] IETF, , "The dnsop mailing list", October 2013.

[dagon-malware]
Dagon, D., "Corrupted DNS Resolution Paths: The Rise of a Malicious Resolution Authority", 2007.

[dns-footprint]
Stoner, E., "DNS footprint of malware", October 2010.

[darkreading-dns]
Lemos, R., "Got Malware? Three Signs Revealed In DNS Traffic", May 2013.

[dnschanger]
Wikipedia, , "DNSchanger", November 2011.

[dnscrypt]
Denis, F., "DNSCrypt", .

[dnscurve]
Bernstein, D., "DNScurve", .

[packetq] , "PacketQ, a simple tool to make SQL-queries against PCAP-files", 2011.

[dnsmezzo]
Bortzmeyer, S., "DNSmezzo", 2009.

[prism] NSA, , "PRISM", 2007.

[crime] Rizzo, J. and T. Dong, "The CRIME attack against TLS", 2012.

[ditl] , "A Day in the Life of the Internet (DITL)", 2002.

[data-protection-directive]
, "European directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data", November 1995.

[passive-dns]
Weimer, F., "Passive DNS Replication", April 2005.

[tor-leak]
, "DNS leaks in Tor", 2013.

Author's Address

Stephane Bortzmeyer
AFNIC
Immeuble International
Saint-Quentin-en-Yvelines 78181
France

Phone: +33 1 39 30 83 46
Email: bortzmeyer+ietf@nic.fr
URI: http://www.afnic.fr/