Technical Analysis of the EDPB Letter to BEREC

ICANN Office of the Chief Technology Officer

OCTO-009
23 April 2020

TABLE OF CONTENTS

This document is part of the OCTO document series. Please see here for a list of documents in the series. If you have questions or suggestions on any of these documents, please send them to octo@icann.org.

This document is published by the Internet Corporation for Assigned Names and Numbers (ICANN). ICANN is a not-for-profit public-benefit corporation that, on behalf of the Internet community, oversees the technical coordination of the top-most level of the Internet's Domain Name System (DNS), and especially its security, stability and resiliency. The authors of this brief are technologists employed by ICANN who have studied and written extensively on issues of Internet routing, the DNS, e-mail, cybersecurity, Internet Protocol (IP) addressing, and other internetworking technologies.

1 Introduction

On 3 December 2019, the European Data Protection Board ("EDPB") sent a letter[1] to the Body of European Regulators for Electronic Communications ("BEREC") responding to questions raised by BEREC concerning the EDPB's interpretation of several key technical elements of the processing of electronic communications data and metadata within the legal framework of Regulation (EU) 2015/2120 concerning open internet access ("OIR"),[2] Regulation (EU) 2016/679 concerning GDPR ("GDPR"),[3] and Directive 2002/58/EC, the ePrivacy Directive ("ePD").[4]

In their letter, the EDPB states:

> *"... the Board considers that processing of data such as the domain name and URL by internet access services providers for traffic management and billing purposes is unlawful, unless consent of all users is obtained."*

ICANN strongly disagrees with this interpretation. If regulators embraced this interpretation, some fundamental functionality all Internet Service Providers (ISPs) must use to provide basic services that all consumers expect would be disallowed.

This document provides an analysis of the potential technical implications of the EDPB's position. Our analysis focuses primarily on potential misinterpretations about how the underlying Internet infrastructure works today. We explain how the processing of DNS data and URLs in traffic data without consent of all end users is necessary for the Internet to function.

2 DNS Data is Traffic Data used to Identify the Source and Destination of a Communication

DNS data such as domain names themselves and the results of the DNS "resolution" process as well as uniform resource locators (URLs) are all electronic communications traffic data. They

1 The letter is published at https://edpb.europa.eu/our-work-tools/our-documents/letters/edpb-response-berec-request-guidance-revision-its-guidelines_en

2 See https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32015R2120

3 See https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679

4 See https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02002L0058-20091219

are part of the information transmitted for the purposes of exchanging information; specifically they are data used to identify the source and destination of a communication.

Computers exchange data with other computers via the Internet Protocol (IP) and identify the source from which packets originate and the destination to which packets are sent based (at least) on IP addresses. Additional information is frequently necessary to identify the final destination, e.g., the point in time when the communication occurred in cases where IP addresses are reassigned dynamically or a port (or service) identifier in cases where Network Address Translation is being used.

Humans generally don't use IP addresses to identify destinations because we typically don't memorize long strings of numbers. Instead, we use the DNS to translate semantic names associated with the destination (more easily learned and remembered by humans) to the IP addresses the computers need to interconnect.

To obtain information on particular domain names, the DNS uses its own metadata to temporarily "cache" responses to queries made by network elements known as "resolvers", typically operated by Internet Service Providers, sent to servers that publish DNS data typically operated by third parties. The cached DNS metadata ensures that when the same query is made by clients of the resolver in the future, the answer can be provided much more quickly, i.e., directly from a local data store (the cache) instead of needing to query potentially multiple remote servers in response to every translation request. Given the scale at which the Internet operates and the criticality of minimizing the time to translate domain names into addresses, caching is an essential function of the operation of the DNS.

3 Intermediary Networks Process DNS Data

Internetworking generally requires machine-to-machine communication across multiple, independently owned and operated networks in a manner generally undetectable to humans. As in the Internet every network is not directly connected to every other network, various network elements, including telecommunication links, routers, etc., are designed to get packets from their source to their destination according to policies set by the administrators of those network elements. These policies include constraints like "fastest", "cheapest", "stay within a single network", etc. Some policies can be complex, based on metadata or packet content, e.g., "packets involved in a communication flow between these domains must not traverse this network" or "traffic coming from this URL must be blocked due to law." All Internet communications are packetized, with each packet being independently forwarded by network elements according to those policies towards the packet's destination. Packets, therefore, often have to traverse multiple intermediate networks and devices, potentially residing in multiple legal jurisdictions, to get from their source to their destination. Unlike some overly-simplified models of how the telephony or other communications networks work, neither the end user who owns the data nor the service providers who process and deliver the data can be certain which of the myriad networks or devices the packets will need to traverse. It is, therefore, not feasible for an intermediary network operator to obtain end user consent in order to process the electronic communication.

The speed, flexibility, scalability, and the resilience of the Internet are due in large part to the ability of network elements to make deterministic calculations to decide which of many paths to use to forward packets. Some of the data fed into those calculations includes various forms of metadata, which we believe the ePD allows. Without this statutory acknowledgement, which the EDPB statement could be interpreted to rescind, the Internet would be unable to function at a global scale.

4 Non-Invasive Monitoring is Necessary to Convey Electronic Communications

In the Annex of its letter, the EDPB refers to the Judgement of the Court of Justice ("ECJ") in the case *Scarlet Extended*[5] in a discussion about traffic monitoring and subsequent filtering. We feel it is important to differentiate the facts of *Scarlet Extended* from technologies that use traffic data monitoring to perform non-invasive filtering.

In *Scarlet Extended*, the contested filtering system for ISPs to combat Internet piracy would have been excessive. As the ECJ held, such a filter would have involved, "*a systematic analysis of all content and the collection and identification of users' IP addresses"*[6]. This would be an invasive filter to protect the individual interests of certain copyright holders. Such a filter would not, however, increase the security or quality of electronic communications.

In contrast, many ISPs regularly perform much less invasive traffic data monitoring for the purposes of filtering based on IP address, domain name, and URL data, with the goal of increasing the secure and effective conveyance of Internet-based communication. That is the purpose of many common Internet management technologies, including common network elements like firewalls, load balancers, traffic scrubbers, etc. As an illustrative example of how these technologies are necessary for the effective conveyance of electronic communications, consider that services and devices on networks are regularly attacked. One of the most common types of attacks, which is also one of the most difficult to mitigate, is a denial of service (DoS) attack. In a DoS attack, which is very easy to implement, an attacker harnesses many data sources, e.g., thousands and sometimes millions of compromised computing devices, to overload a remote network's capacity and/or capabilities. To both protect against the threat of DoS attacks and to thwart those attacks when they're in progress, network managers use monitoring technologies to watch for attack patterns across their networks. When an attack is identified, technical reconfigurations can be quickly made to isolate, filter, and eventually dissipate the DoS traffic. Such filtering defenses require active monitoring and analysis of electronic communications metadata. Disallowing such monitoring would therefore immediately threaten the security and stability of the Internet.

Such monitoring is therefore both "necessary for the conveyance of a communication" (Article 5 (1) of the ePD) and "facilitating the transmission of a communication over an electronic communications network" (Article 5 (3) of the ePD). There is a risk that the EDPB's letter might be understood as stating that only the minimum of what is technically required to convey a communication *at all* would be covered by the aforementioned clauses. Such a narrow

5 ECJ, Judgment of 24 November 2011, *Scarlet Extended*, C-70/10, EU:C:2011:771, available at: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:62010CJ0070

6 ECJ, Scarlet Extended, para. 51

approach would conflict with the reality of Internet-based services and users' expectations about Internet-based communication. To illustrate this by way of an example from the analogue world: To transport persons on a road from location A to location B in a car, neither seats nor safety belts or headlights would be required in a narrow technical sense. However, for good reasons (in particular: security), users (and law enforcement) would regard these features to be *necessary for the conveyance of a person*. Similarly, the term "necessary for the conveyance of a communication" in Article 5 of the ePD can only refer to all measures that are necessary to not only transport a signal from A to B, but also to do so in a *secure and adequate manner in accordance with the state of the art*.

5 Traffic Management

Putting the legality of modern traffic management technology into question risks reducing the ability of network operators of all kinds from ensuring their networks can cope with traffic surges, e.g., as a result of COVID-19 pandemic movement restrictions resulting in significant increases in the use of the Internet. In this section, we analyze three parts of the EDPB's interpretation related to traffic management.

5.1 Traffic Management does not Violate Privacy Principles

In the Annex of its letter, the EDPB lays out its case for the necessity of balancing privacy regulations with critical traffic management efforts. Of particular note, the EDPB states:

> *"... it is worth recalling that the processing of personal data retained in the context of traffic management must also respect other principles that derive from the GDPR and the ePD."*

The goal of traffic management is to improve the scalability, resilience, and/or speed of electronic communications. It is a data science that takes various inputs, including metadata, and optimizes the network's capabilities as an output. That a natural person may be behind the IP addresses of a device or a port at a certain time and date is not pertinent to such analysis and is not taken into account. The association of network metadata to personal data can only be made by humans or software when extra steps are taken beyond traffic management, and such actions are already well covered by the GDPR and the ePD.

5.2 Domain Names and URLs must be Processed for the Purposes of Traffic Management

In the Annex of its letter, the EDPB states:

> *"IAS service providers do not require information included in the transport layer payload (like domain names or URLs) to convey a communication on an electronic communication network.*
>
> *Therefore, domain names and URLs cannot be considered 'traffic data' as defined in Article 2(b) ePD and they cannot be processed under the provisions in Article 6 ePD.*

> *Article 5(1) ePD allows for the 'technical storage which is necessary for the conveyance of a communication'. However, domain names and URLs are not necessary for IAS service providers to convey a communication.*
>
> *Even the option of basing traffic management on consent is not unconstrained."*

The assertion that IAS providers "*do not require information included in the transport layer payload*" makes assumptions about network and application architecture that do not fully capture how today's Internet operates. While in an ideal world where, for example, cybersecurity attacks would not exist, and where efficient traffic management is unnecessary as a result of unlimited high-speed global network capacities, the various protocol layers would be held sacrosanct and "layering violations" would not occur. In our real world, however, network elements such as routers, switches, network address translators, firewalls, etc., frequently need to "peek" into layers above (and below) the network layer they primarily operate within in order to meet nanosecond-scale performance requirements, thwart imminent attacks, optimize network traffic flows, prevent network abuse, and allow for appropriate cost sharing.

Internet communications are typically initiated by references to domain names. These domain names need to be translated into IP addresses. In both large-scale and mission-critical Internet access services, load balancers look at domain names and IP addresses in URLs, as well as other higher-layer protocol elements, in order to pre-fetch answers to reduce overall network latency. Such optimization schemes are common and are expected by users, however those users are, in the vast majority of cases, unaware of these services.

Additionally, domain names, IP addresses, and URL information are automatically analyzed on the wire by anti-malware processes run at many (possibly most) ISPs. Users are easily tricked by people who are initiating fraud, and ISPs believe it is important to protect users from fraud (and protect their own networks from abuse by unwitting customers) by preemptively detecting and denying malware sites and services by name or URL. This traffic data is a critical component of how these automated protections work.

5.3 Differentiated Services are not an Effective Traffic Management Technique

The EDPB also suggests in the Annex:

> *"IAS provider*[s] *could achieve the objectives of traffic management standardizing and using data available at the IP header like the Explicit Congestion Notification (ECN) or the Differentiated Services Code Point (DSCP). Therefore, the EDPB is of the view that processing of domain names and URLs by providers of IAS is not necessary to conduct traffic management."*

This statement ignores both the limitations of ECN and its lack of universality. ECN is designed to be helpful in only some, not all, congestion scenarios. It is only applied against best-effort packets and not to any high priority traffic. Importantly, for ECN to be effective as a traffic management tool for an IAS provider in these limited scenarios, it requires both the next hop downstream of the IAS provider and the next hop upstream of the IAS provider to have ECN

enabled. But this is not often the case. A 2019 research paper[7] reviewed all of the previous studies of the status of ECN deployment and then conducted new studies on ECN deployment., and confirmed that even though ECN was standardized in 2001, and even though web servers enjoy wide ECN support, the enabling of ECN on switches at ISPs remains very low. This means that the techniques suggested by the EDPB are not equally, and not even similarly, effective alternatives. Consequently, the processing of domain names and URLS, as illustrated above, is necessary for the conveyance of communication.

As a consequence of the statement by the EDPB, these techniques, which improve and protect the network and its services as well as the reduce risk to end users making use of those networks and services, would likely need to be viewed as illegal. In addition to potentially reducing the scalability, resilience, and performance of the Internet, this statement could have the ironic effect of enabling attackers to make use of the legal system to increase the vulnerability of IAS providers and their customers.

6 Conclusion

In many, if not all, technical fields such as Internet engineering, one common challenge is presenting information in such a way as to allow individuals without a technical background to understand the information. In many cases, highly technical information is presented in a vastly simplified form, with caveats indicating the information is not an accurate representation of how the technologies actually work. Unfortunately, it is not uncommon for those caveats to be ignored and for rules, policies, regulations, and legislation to be based on the vastly simplified representations of the technology. The resulting constraints, while often well intentioned, can have very negative effects on operators of the technologies, reducing the resiliency, scalability, and performance of their networks as well as increasing costs, which are invariably passed on to end users.

ICANN's mission includes promoting best practices that enable a more secure DNS ecosystem. We trust we have shown in our analysis how important it is that the European Union legislation and European Regulators allow the best practices described in our analysis that are widely prevalent in ISPs today. A secure DNS ecosystem fundamentally requires these activities to occur on a constant basis to ensure that bad actors who seek to harm Internet users do not succeed. We are therefore concerned about the potential but serious side effects that may occur should the aforementioned technical consequences not be adequately considered.

[7] See https://www.jstage.jst.go.jp/article/transinf/E102.D/5/E102.D_2018NTP0006/_article/-char/en