

Network Working Group
Internet-Draft
Intended status: Informational
Expires: 24 April 2025

L. Pardue
Cloudflare
21 October 2024

Maintaining Protocols Using Grease and Variability
draft-edm-protocol-greasing-04

Abstract

Long-term interoperability of protocols is an important goal of the network standards process. Deployment success can depend on supporting change, which can include modifying how the protocol is used, extending the protocol, or replacing the protocol. This document presents concepts, considerations, and techniques related to protocol maintenance, such as greasing or variability. The intended audience is protocol designers and implementers.

About This Document

This note is to be removed before publishing as an RFC.

The latest revision of this draft can be found at <https://intarchboard.github.io/draft-protocol-greasing/draft-edm-protocol-greasing.html>. Status information for this document may be found at <https://datatracker.ietf.org/doc/draft-edm-protocol-greasing/>.

Source for this draft and an issue tracker can be found at <https://github.com/intarchboard/draft-protocol-greasing>.

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 <https://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 24 April 2025.

Copyright Notice

Copyright (c) 2024 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

Table of Contents

| | |
|---|---|
| 1. Introduction | 2 |
| 2. Conventions and Definitions | 3 |
| 3. Considerations for Applying Greasing | 3 |
| 3.1. Don't Handle Grease Values as a Special Case | 4 |
| 3.2. Use Unpredictable Grease Values | 4 |
| 3.3. Use Grease Values at Unpredictable Times | 4 |
| 3.4. Define and Register Grease Value Ranges | 4 |
| 3.4.1. Effectively Instructing IANA about grease | 5 |
| 4. Considerations for Increasing Protocol Variability | 5 |
| 4.1. Example: QUIC frames | 5 |
| 5. Considerations for Protocol Versions | 6 |
| 6. Security Considerations | 6 |
| 7. IANA Considerations | 6 |
| 8. References | 6 |
| 8.1. Normative References | 6 |
| 8.2. Informative References | 7 |
| Acknowledgments | 8 |
| Author's Address | 8 |

1. Introduction

Long-term interoperability of protocols is an important goal of the network standards process [MAINTENANCE]. Protocol deployment success [SUCCESS] can depend on supporting change, which can include modifying how the protocol is used, extending the protocol, or replacing the protocol.

Greasing, a technique initially designed for TLS [GREASE] and later adopted by other protocols such as QUIC [QUIC], can help support the long-term viability of protocol extension points. In these protocols, extension codepoints are reserved only for greasing and when received must be ignored. Greasing is suitable for many protocols but not all; Section 3.3 of [VIABILITY] discusses the applicability and limitations of greasing. Section 3 provides additional protocol maintenance considerations.

Applications are built with the intent of serving user needs [END-USERS], which might only require support for a subset of protocol features. Adapting to changing user needs is a maintenance activity. For example, an HTTP deployment focused on downloads might want to add support for uploads. Changing use of the application and transport protocol features can affect the deployment's network traffic profile. If expectations have been formed around historical patterns of use i.e., ossification, introducing change might lead to deployment problems. Section 4 presents considerations about how intentionally increasing the variability of protocols can mitigate some of these concerns.

Protocol extensions can provide longevity in the face of changing needs or environment. However, a replacement protocol might be preferred when extensions are not adequate or feasible. A protocol replacement could aggregate common extensions and possibly make them mandatory, effectively defining a new baseline that can simplify deployment and interoperability. A replacement protocol version may or may not be compatible with other versions. A protocol may or may not have a mechanism for version selection or agility. Section 5 presents considerations about designing for and/or implementing version negotiation and migration.

2. Conventions and Definitions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Considerations for Applying Greasing

Greasing can take many forms, depending on the protocol and the nature of its extension points.

Many protocols register values, codepoints, or numbers in a limited space. A common approach that has developed in more recent protocols is to reserve a subset of the space for greasing (see [GREASE], Section 18.1 of [QUIC], or Section 7.2.8 of [HTTP/3]). Values reserved for the purpose of greasing are herein referred to as grease values. Implementations that receive grease values are required to ignore them. More background to this approach is given in Section 3.3 of [VIABILITY]. This section provides concrete suggestions for its usage.

3.1. Don't Handle Grease Values as a Special Case

It is assumed that endpoints should implement robust and broad extension handling. A receiver or a parser implementation should not treat grease values as individually special. Instead of identifying each grease value explicitly, it is better to have a "catch all" mechanism that can handle receipt of unknown extensions, whether grease values or not, gracefully or without error.

3.2. Use Unpredictable Grease Values

It is recommended that senders pick an unpredictable grease value to include in relevant protocol elements. This ensures that receiver greasing requirements are exercised. Using predictable grease values risks ossification. To increase the variety of grease values, it is advised to reserve a large range. However, the specific size and distribution of the grease range needs to accommodate the protocol constraints. For instance, protocols that use 8-bit fields may find it too costly to dedicate many grease values, while 32-bit or 64-bit fields are likely to have no limitations.

3.3. Use Grease Values at Unpredictable Times

It is recommended that senders use grease values at unpredictable times or sequence points during protocol interactions. This avoids receivers unintentionally ossifying on the occurrence of greasing in the temporal or spatial domain.

3.4. Define and Register Grease Value Ranges

It is recommended that large grease value sets are allocated in protocol documents by defining a unique algorithm, to increase the chance that receiver greasing requirements are exercised. However, the choice of algorithm needs to consider the spread of values and the size of contiguous blocks between grease values. It is common for protocol extension designers to want to reserve a contiguous block of code points in order to aid iteration and experimentation. Small contiguous blocks increase the chance that such reservations might unintentionally use grease values, which could lead to interoperability failures.

3.4.1. Effectively Instructing IANA about grease

Protocol designers might ask IANA to create new registries for their extension points. When greasing, it is recommended that only a single entry for the entire grease value set is registered. When an algorithm has been used, it should be included in the entry; see for example <https://www.iana.org/assignments/http3-parameters/http3-parameters.xhtml#http3-parameters-frame-types>.

Grease values must not be used or registered for any other purpose. Registries should include a label to identify the protected grease value range; a label of "reserved" may be confused with other ranges that are reserved for private or experimental extensions. An implementer that conflates these two meanings may cause a new class of interoperability failure. Therefore a label such as "reserved for greasing" may help to avoid the confusion.

4. Considerations for Increasing Protocol Variability

Greasing can maintain protocol extensibility by falsifying active use of its extension points. However, greasing alone does not ensure positive use of extension mechanisms. A protocol may define a wide-ranging extension capability that remains unused in the absence of real use cases. This can lead to ossification that does not expect extensions, leading to interoperability problems later on.

Long-term maintenance and interoperability can be ensured by exercising extension points positively. To some extent this can be thought of as protocol fuzzing. This might be difficult to exercise because varying the protocol elements might change the outcome of interactions, leading to real errors. However, some protocols allow elements to be safely changed, as shown in the following examples.

4.1. Example: QUIC frames

QUIC packets contain frames. Receivers might build expectations on the longitudinal aspects of packets or frames - size, ordering, frequency, etc. A sender can quite often manipulate these parameters and stay compliant to the requirements of the QUIC protocol.

A QUIC stream is an ordered reliable byte stream that is serialized as a sequence of STREAM frames with a length and offset. Receivers are expected to reassemble frames, which could arrive in any order, into an ordered reliable byte stream that is readable by applications.

A form of positive testing is for a sender to unpredictably order the STREAM frames that it transmits. For example, varying the sequence order of offset values. This allows to exercise the QUIC reassembly features of the receiver with the expectation that no failure would occur. However, doing this may introduce delay or stream head-of-line blocking that affects the performance aspects of a transmission, which may not be acceptable for a given use case. As such, positive testing might be most appropriate to use in a subset of connections, or phases within a connection.

5. Considerations for Protocol Versions

There are intrinsic and well-documented issues related to testing version negotiation of protocols; see [EXTENSIBILITY] and Sections 2.1 and 3.2 of [VIABILITY]. This section will be expanded with advice for protocol designers and implementers about how to approach these problems.

6. Security Considerations

The considerations in [MAINTENANCE], [GREASE], [END-USERS], and [VIABILITY] all apply to the topics discussed in this document.

The use of protocol features, extensions, and versions can already allow fingerprinting [PRIVCON]. Any techniques that change parameters in any way, including but not limited to those discussed in this document, can affect fingerprinting. A deeper analysis of this topic has been deemed out of scope.

7. IANA Considerations

This document has no IANA actions itself. Guidance on how other documents can effectively instruct IANA about protocol greasing is provided in Section 3.4.1

8. References

8.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/rfc/rfc2119>>.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/rfc/rfc8174>>.

8.2. Informative References

[END-USERS]

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

[EXTENSIBILITY]

Carpenter, B., Aboba, B., Ed., and S. Cheshire, "Design Considerations for Protocol Extensions", RFC 6709, DOI 10.17487/RFC6709, September 2012, <<https://www.rfc-editor.org/rfc/rfc6709>>.

[GREASE]

Benjamin, D., "Applying Generate Random Extensions And Sustain Extensibility (GREASE) to TLS Extensibility", RFC 8701, DOI 10.17487/RFC8701, January 2020, <<https://www.rfc-editor.org/rfc/rfc8701>>.

[HTTP/3]

Bishop, M., Ed., "HTTP/3", RFC 9114, DOI 10.17487/RFC9114, June 2022, <<https://www.rfc-editor.org/rfc/rfc9114>>.

[MAINTENANCE]

Thomson, M. and D. Schinazi, "Maintaining Robust Protocols", RFC 9413, DOI 10.17487/RFC9413, June 2023, <<https://www.rfc-editor.org/rfc/rfc9413>>.

[PRIVCON]

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/rfc/rfc6973>>.

[QUIC]

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/rfc/rfc9000>>.

[SUCCESS]

Thaler, D. and B. Aboba, "What Makes for a Successful Protocol?", RFC 5218, DOI 10.17487/RFC5218, July 2008, <<https://www.rfc-editor.org/rfc/rfc5218>>.

[VIABILITY]

Thomson, M. and T. Pauly, "Long-Term Viability of Protocol Extension Mechanisms", RFC 9170, DOI 10.17487/RFC9170, December 2021, <<https://www.rfc-editor.org/rfc/rfc9170>>.

Acknowledgments

This work is a summary of the topics discussed during EDM meetings.
The contributors at those meetings are thanked.

Author's Address

Lucas Pardue
Cloudflare
Email: lucas@lucaspardue.com

Network Working Group
Internet-Draft
Intended status: Informational
Expires: 8 May 2025

M. Nottingham
Cloudflare
S. Krishnan
Cisco
4 November 2024

IAB AI-CONTROL Workshop Report
draft-iab-ai-control-report-00

Abstract

The AI-CONTROL Workshop was convened by the Internet Architecture Board (IAB) in September 2024. This report summarizes its significant points of discussion and identifies topics that may warrant further consideration and work.

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 <https://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 8 May 2025.

Copyright Notice

Copyright (c) 2024 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

Table of Contents

| | |
|--|----|
| 1. Introduction | 2 |
| 1.1. Chatham House Rule | 3 |
| 1.2. Views Expressed in this Report | 3 |
| 2. Overview of the AI Crawling Landscape | 3 |
| 2.1. Crawl Time vs. Inference Time | 4 |
| 2.1.1. Multiple Uses for Crawl Data | 4 |
| 2.1.2. Application of Preferences | 5 |
| 2.2. Trust | 5 |
| 2.3. Attachment | 6 |
| 2.3.1. robots.txt (and similar) | 6 |
| 2.3.2. Embedding | 6 |
| 2.3.3. Registries | 7 |
| 2.4. Vocabulary | 8 |
| 3. Conclusions | 8 |
| 3.1. Potential Standards Work | 8 |
| 3.1.1. Out of Initial Scope | 9 |
| 4. Security Considerations | 9 |
| 5. Informative References | 9 |
| Appendix A. About the Workshop | 9 |
| A.1. Agenda | 10 |
| A.1.1. Thursday 2024-09-19 | 10 |
| A.1.2. Friday 2024-09-20 | 10 |
| A.2. Attendees | 10 |
| IAB Members at the Time of Approval | 11 |
| Acknowledgements | 12 |
| Authors' Addresses | 12 |

1. Introduction

The Internet Architecture Board (IAB) holds occasional workshops designed to consider long-term issues and strategies for the Internet, and to suggest future directions for the Internet architecture. This long-term planning function of the IAB is complementary to the ongoing engineering efforts performed by working groups of the Internet Engineering Task Force (IETF).

The Internet is one of the major sources of data used to train large language models (LLMs, or more generally "AI"). Because this use was not envisioned by most publishers of information on the Internet, a means of expressing the owners' preferences regarding AI crawling has emerged, sometimes backed by law (e.g., in the European Union's AI Act [AI-ACT]).

The IAB convened the AI-CONTROL Workshop to "explore practical opt-out mechanisms for AI and build an understanding of use cases, requirements, and other considerations in this space." [CFP] In

particular, the emerging practice of using the Robots Exclusion Protocol [RFC9309] -- also known as "robots.txt" -- has been uncoordinated, and may or may not be a suitable way to control AI crawlers. However, discussion was not limited to consideration of robots.txt, and approaches other than opt-out were considered.

To ensure many viewpoints were represented, the program committee invited a broad selection of technical experts, AI vendors, content publishers, civil society advocates, and policymakers.

1.1. Chatham House Rule

Participants agreed to conduct the workshop under the Chatham House Rule [CHATHAM-HOUSE], so this report does not attribute statements to individuals or organizations without express permission. Most submissions to the workshop were public and thus attributable; they are used here to provide substance and context.

Appendix A.2 lists the workshop participants, unless they requested that this information be withheld.

1.2. Views Expressed in this Report

This document is a report on the proceedings of the workshop. The views and positions documented in this report are expressed during the workshop by participants and do not necessarily reflect IAB's views and positions.

Furthermore, the content of the report comes from presentations given by workshop participants and notes taken during the discussions, without interpretation or validation. Thus, the content of this report follows the flow and dialogue of the workshop but does not attempt to capture a consensus.

2. Overview of the AI Crawling Landscape

The workshop began by surveying the state of AI control.

Currently, Internet publishers express their preferences for how their content is treated for purposes of AI training using a variety of mechanisms, including declarative ones, such as terms of service and robots.txt [RFC9309], and active ones, such as use of paywalls and selective blocking of crawlers (e.g., by IP address, User-Agent).

There was disagreement about the implications of AI opt-out overall. Research indicates that the use of such controls is becoming more prevalent, reducing the availability of data for AI training. She participants expressed concern about the implications of this --

although at least one AI vendor seemed less concerned by this, indicating that "there are plenty of tokens available" for training, even if many opt out. Others expressed a need to opt out of AI training because of how they perceive its effects on their control over content, seeing AI as usurping their relationships with customers and a potential threat to whole industries.

However, there was quick agreement that both viewpoints were harmed by the current state of AI opt-out -- a situation where "no one is better off" (in the words of one participant).

Much of that dysfunction was attributed to the lack of coordination and standards for AI opt out. Currently, content publishers need to consult with each AI vendor to understand how to opt out of training their products, as there is significant variance in each vendor's behaviour. Furthermore, publishers need to continually monitor both for new vendors, and for changes to the policies of the vendors they are aware of.

Underlying those immediate issues, however, are significant constraints that could be attributed to uncertainties in the legal context, the nature of AI, and the implications of needing to opt out of crawling for it.

2.1. Crawl Time vs. Inference Time

Perhaps most significant is the "crawl time vs. inference time" problem. Statements of preference are apparent at crawl time, bound to content either by location (e.g. robots.txt) or embedded inside the content itself as metadata. However, the target of those directives is often disassociated from the crawler, either because the crawl data is not only used for training AI models, or because the preferences are applicable at inference time.

2.1.1. Multiple Uses for Crawl Data

A crawl's data might have multiple uses because the vendor also has another product that uses it (e.g., a search engine), or because the crawl is performed by a party other than the AI vendor. Both are very common patterns: operators of many Internet search engines also train AI models, and many AI models use third party crawl data. In either case, conflating different uses can change the incentives for publishers to cooperate with the crawler.

Well-established uses of crawling such as Internet search were seen by participants as at least partially aligned with the interests of publishers: they allow their sites to be crawled, and in return they receive higher traffic and attention due to being in the search

index. However, several participants pointed out that this symbiotic relationship does not exist for AI training uses -- with some viewing AI as hostile to publishers, because it has the capacity to take traffic away from their sites.

Therefore, when a crawler has multiple uses that include AI, participants observed that "collateral damage" was likely for non-AI uses, especially when publishers take more active control measures such as blocking or paywalls to protect their interests.

Several participants expressed concerns about this phenomenon's effects on the ecosystem, effectively "locking down the Web" with one opining that there were implications on freedom of expression overall.

2.1.2. Application of Preferences

When data is used to train an LLM, the resulting model does not have the ability to only selectively use a portion of it when performing a task, because inference uses the whole model, and it is not possible to identify specific input data for its use in doing so.

This means that while publishers preferences may be available when content is crawled, they generally are not when inference takes place. Those preferences that are stated in reference to use by AI -- for example, "no military uses" or "non-commercial only" cannot be applied by a general-purpose "foundation" model.

This leaves a few unappealing choices to AI vendors that wish to comply with those preferences. They can simply omit such data from foundation models, thereby reducing their viability. Or, they can create a separate model for each permutation of preferences -- with a likely proliferation of models as the set of permutations expands.

Compounding this issue was the observation that preferences change over time, whereas LLMs are created over long time frames and cannot easily be updated to reflect those changes. Of particular concern to some was how an opt-out regime makes the default stickier.

2.2. Trust

This disconnection between the statement of preferences and its application was felt by participants to contribute to a lack of trust in the ecosystem, along with the typical lack of attribution for data sources in LLMs, lack of an incentive for publishers to contribute data, and finally (and most noted) a lack of any means of monitoring compliance with preferences.

This lack of trust led some participations to question whether communicating preferences is sufficient in all cases without an accompanying way to mitigate or track cases of those preferences being followed. Some participants also indicated that lack of trust was the primary cause of increasingly prevalent blocking of AI crawler IP addresses, among other measures.

2.3. Attachment

One of the primary focuses of the workshop was on `_attachment_` -- how preferences are associated with content on the Internet. A range of mechanisms was discussed.

2.3.1. robots.txt (and similar)

The Robots Exclusion Protocol [RFC9309] is widely recognised by AI vendors as an attachment mechanism for preferences. Several deficiencies were discussed.

First, it does not scale to offer granular control over large sites where authors might want to express different policies for a range of content (for example, YouTube).

Robots.txt also is typically under the control of the site administrator. If a site has content from many creators (as is often the case for social media and similar platforms), the administrator may not allow them to express their preferences fully, or at all.

If content is copied or moved to a different site, the preferences at the new site need to be explicitly transferred, because robots.txt is a separate resource.

These deficiencies led many participants to feel that robots.txt cannot be the only solution to opt-out: rather, it should be part of a larger system that addresses its shortcomings.

Participants noted that other, similar attachment mechanisms have been proposed. However, none appear to have gained as much attention or implementation (both by AI vendors and content owners) as robots.txt.

2.3.2. Embedding

Another mechanism for associating preferences with content is to embed them into the content itself. Many formats used on the Internet allow this; for example, HTML has the `<meta>` tag, images have XMP and similar metadata sections, and XML and JSON have rich potential for extensions to carry such data.

Embedded preferences were seen to have the advantage of granularity, and of "travelling with" content as it is produced, when it is moved from site to site, or when it is stored offline.

However, several participants pointed out that embedded preferences are easily stripped from most formats. This is a common practice for reducing the size of a file (thereby improving performance when downloading it), and for assuring privacy (since metadata often leaks information unintentionally).

Furthermore, some types of content are not suitable for embedding. For example, it is not possible to embed preferences into purely textual content, and Web pages with content from several producers (such as a social media or comments feed) cannot easily reflect preferences for each one.

Participants noted that the means of embedding preferences in many formats would need to be determined by or coordinated with organisations outside the IETF. For example, HTML and many image formats are maintained by external bodies.

2.3.3. Registries

In some existing copyright management regimes, it is already common to have a registry of works that is consulted upon use. For example, this approach is often used for photographs, music, and video.

Typically, registries use hashing mechanisms to create a "fingerprint" for the content that is robust to changes.

Using a registry decouples the content in question from its location, so that it can be found even if moved. It is also claimed to be robust against stripping of embedded metadata, which is a common practice to improve performance and/or privacy.

However, several participants pointed out issues with deploying registries at Internet scale. While they may be effective for (relatively) closed and well-known ecosystems such as commercial music publishing, applying them to a diverse and very large ecosystem like the Internet has proven problematic.

2.4. Vocabulary

Another major focus area for the workshop was on `_vocabulary_` -- the specific semantics of the opt-out signal. Several participants noted that there are already many proposals for vocabularies, as well as many conflicting vocabularies already in use. Several examples were discussed, including where existing terms were ambiguous, did not address common use cases, or were used in conflicting way by different actors.

Although no conclusions regarding exact vocabulary were reached, it was generally agreed that a complex vocabulary is unlikely to succeed.

3. Conclusions

Participants seemed to agree that on its current path, the ecosystem is not sustainable. As one remarked, "robots.txt is broken and we broke it."

Legal uncertainty, along with fundamental limitations of opt-out regimes pointed out above, limit the effectiveness of any technical solution, which will be operating in a system unlike either robots.txt (where there is a symbiotic relationship between content owners and the crawlers) or copyright (where the default is effectively opt-in, not opt-out).

However, the workshop ended with general agreement that positive steps could be taken to improve communication of preferences from content owners for AI use cases. In discussion, it was evident that discovery of preferences from multiple attachment mechanisms is necessary to meet the diverse needs of content authors, and that therefore defining how they are combined is important.

We outline a proposed standard program below.

3.1. Potential Standards Work

The following items were felt to be good starting points for IETF work:

- * Attachment to Web sites by location (in robots.txt or a similar mechanism)
- * Attachment via embedding in IETF-controlled formats (e.g., HTTP headers)
- * Definition of a common core vocabulary
- * Definition of the overall regime; e.g., how to combine preferences discovered from multiple attachment mechanisms

It would be expected that the IETF would coordinate with other SDOs to define embedding in other formats (e.g., HTML).

3.1.1. Out of Initial Scope

It was broadly agreed that it would not be useful to work on the following items, at least to begin with:

- * Enforcement mechanisms for preferences
- * Registry-based solutions
- * Identifying or authenticating crawlers and/or content owners
- * Audit or transparency mechanisms

4. Security Considerations

TODO

5. Informative References

[CHATHAM-HOUSE]

Chatham House, "Chatham House Rule", n.d.,
<<https://www.chathamhouse.org/about-us/chatham-house-rule>>.

[CFP] Internet Architecture Board, "IAB Workshop on AI-CONTROL", n.d.,
<<https://datatracker.ietf.org/group/aicontrolws/about/>>.

[PAPERS] Internet Architecture Board, "IAB Workshop on AI-CONTROL Materials", n.d.,
<<https://datatracker.ietf.org/group/aicontrolws/materials/>>.

[AI-ACT] European Parliament, "Regulation (eu) 2024/1689 of the European Parliament and of the Council", 13 June 2024,
<<https://eur-lex.europa.eu/eli/reg/2024/1689/oj>>.

[RFC9309] Koster, M., Illyes, G., Zeller, H., and L. Sassman, "Robots Exclusion Protocol", RFC 9309, DOI 10.17487/RFC9309, September 2022,
<<https://www.rfc-editor.org/rfc/rfc9309>>.

Appendix A. About the Workshop

The AI-CONTROL Workshop was held on 2024-09-19 and 2024-09-29 at Wilkinson Barker Knauer in Washington DC, USA.

Workshop attendees were asked to submit position papers. These papers are published on the IAB website [PAPERS], unless the submitter requested it be withheld.

The workshop was conducted under the Chatham House Rule [CHATHAM-HOUSE], meaning that statements cannot be attributed to individuals or organizations without explicit authorization.

A.1. Agenda

This section outlines the broad areas of discussion on each day.

A.1.1. Thursday 2024-09-19

Setting the stage An overview of the current state of AI opt-out, its impact, and existing work in this space

Lightning talks A variety of perspectives from participants

A.1.2. Friday 2024-09-20

Opt-Out Attachment: robots.txt and beyond Considerations in how preferences are attached to content on the Internet

Vocabulary: what opt-out means What information the opt-out signal needs to convey

Discussion and wrap-up Synthesis of the workshop's topics and how future work might unfold

A.2. Attendees

Attendees of the workshop are listed with their primary affiliation. Attendees from the program committee (PC) and the Internet Architecture Board (IAB) are also marked.

- * Jari Arkko, Ericsson
- * Hirochika Asai, Preferred Networks
- * Farzaneh Badiei, Digital Medusa (PC)
- * Fabrice Canel, Microsoft (PC)
- * Lena Cohen, EFF
- * Alissa Cooper, Knight-Georgetown Institute (PC, IAB)
- * Marwan Fayed, Cloudflare
- * Christopher Flammang, Elsevier
- * Carl Gahnberg
- * Max Gendler, The News Corporation
- * Ted Hardie
- * Dominique Hazaeï\2101-Massieux, W3C

- * Gary Ilyes, Google (PC)
- * Sarah Jennings, UK Department for Science, Innovation and Technology
- * Paul Keller, Open Future
- * Elizabeth Kendall, Meta
- * Suresh Krishnan, Cisco (PC, IAB)
- * Mirja Köhlerwind, Ericsson (PC, IAB)
- * Greg Leppert, Berkman Klein Center
- * Greg Lindahl, Common Crawl Foundation
- * Mike Linksvayer, GitHub
- * Fred von Lohmann, OpenAI
- * Shayne Longpre, Data Provenance Initiative
- * Don Marti, Raptive
- * Sarah McKenna, Alliance for Responsible Data Collection; CEO, Sequentum
- * Eric Null, Center for Democracy and Technology
- * Chris Needham, BBC
- * Mark Nottingham, Cloudflare (PC)
- * Paul Ohm, Georgetown Law (PC)
- * Braxton Perkins, NBC Universal
- * Chris Petrillo, Wikimedia
- * Sebastian Posth, Liccium
- * Michael Prorock
- * Matt Rogerson, Financial Times
- * Peter Santhanam, IBM
- * Jeffrey Sedlik, IPTC/PLUS
- * Rony Shalit, Alliance For Responsible Data Collection; Bright Data
- * Ian Sohl, OpenAI
- * Martin Thomson, Mozilla
- * Thom Vaughan, Common Crawl Foundation (PC)
- * Kat Walsh, Creative Commons
- * James Whymark, Meta

The following participants requested that their identity and/or affiliation not be revealed:

- * A government official

IAB Members at the Time of Approval

Internet Architecture Board members at the time this document was approved for publication were:

TBC

Acknowledgements

The Program Committee and the IAB would like to thank Wilkinson Barker Knauer for their generosity in hosting the workshop.

We also thank our scribes for capturing notes that assisted in production of this report:

- * Zander Arnao
- * Andrea Dean
- * Patrick Yurky

Authors' Addresses

Mark Nottingham
Cloudflare
Pahran
Australia
Email: mnot@mnot.net
URI: <https://www.mnot.net/>

Suresh Krishnan
Cisco
Email: suresh.krishnan@gmail.com

Network Working Group
Internet-Draft
Intended status: Informational
Expires: 22 December 2024

M. Kählewind
D. Dhody
M. Knodel
20 June 2024

IAB Barriers to Internet Access of Services (BIAS) Workshop Report
draft-iab-bias-workshop-report-02

Abstract

The "Barriers for Internet Access of Services (BIAS)" workshop was convened by the Internet Architecture Board (IAB) from January 15-17, 2024 as a three-day online meeting. Based on the submitted position papers, the workshop covered three areas of interest: the role of community networks in Internet Access of Services; reports and comments on the observed digital divide; and measurements of censorship and censorship circumvention. This report summarizes the workshop's discussion and serves as a reference for reports on the current barriers to Internet Access.

Note that this document is a report on the proceedings of the workshop. The views and positions documented in this report were expressed during the workshop by participants and do not necessarily reflect IAB's views and positions.

About This Document

This note is to be removed before publishing as an RFC.

The latest revision of this draft can be found at <https://intarchboard.github.io/draft-iab-bias-workshop-report/draft-iab-bias-workshop-report.html>. Status information for this document may be found at <https://datatracker.ietf.org/doc/draft-iab-bias-workshop-report/>.

Source for this draft and an issue tracker can be found at <https://github.com/intarchboard/draft-iab-bias-workshop-report>.

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 <https://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 22 December 2024.

Copyright Notice

Copyright (c) 2024 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

Table of Contents

| | | |
|--------|--|----|
| 1. | Introduction | 3 |
| 1.1. | About this workshop report content | 3 |
| 2. | Workshop Scope and Discussion | 4 |
| 2.1. | Session 1: Community Networks - Their Role in Internet Access of Services | 4 |
| 2.1.1. | The Quality of Community Networks | 4 |
| 2.1.2. | Strengthening Community Networks | 5 |
| 2.1.3. | Discussion | 5 |
| 2.2. | Session 2: Digital Divide - Reports and Comments | 6 |
| 2.2.1. | Disparities in service provision | 6 |
| 2.2.2. | Lack of consistent acceptance of language scripts | 7 |
| 2.2.3. | Web Affordability and Inclusiveness | 7 |
| 2.2.4. | Discussion | 7 |
| 2.3. | Session 3: Censorship - Reports and Circumvention | 8 |
| 2.3.1. | Censorship Orders, Measurements, and Device Analysis | 8 |
| 2.3.2. | Use of VPNs for Censorship Circumvents and User Expectations | 10 |
| 2.3.3. | Discussion | 10 |
| 2.4. | Key Take Aways | 10 |
| 3. | Informative References | 11 |
| | Appendix A. Position Papers | 13 |
| | Appendix B. Workshop Participants | 15 |

Appendix C. Workshop Program Committee 15
 IAB Members at the Time of Approval 15
 Acknowledgments 15
 Authors' Addresses 15

1. Introduction

The Internet as part of the critical infrastructure affects many aspects of our society significantly, although it impacts different parts of society differently. The Internet is an important tool to reach the Sustainable Development Goals (SDG) [SDG] and to globally support human rights. Consequently, the lack of meaningful access to digital infrastructure and services is also a form of disenfranchisement.

Solely having Internet access is not enough. At the same time as we work to connect the next billion people and reduce the digital divide, it is also important to understand persistent and novel inequalities in the digital age when accessing content and services. There are more and more barriers to meaningful access to the services and applications that run on the Internet. Even if Internet connectivity is available, information and service access may remain challenged and unequal.

This IAB workshop has aimed

- * to collect reports about barriers to accessing content and services on the Internet, e.g. based on filtering, and blocking as well as due to general inequality of technological capabilities, like device or protocol limitations.
- * to help the Internet community get a better understanding of how the Internet functions in different parts of the world and which technology or techniques need to be used to gain access to content.
- * to build an understanding of what "being connected" to the Internet means: What is the Internet to users? What is needed to be meaningfully connected? What are the minimum requirements to be able to access certain parts of the content and services provided over the Internet?

1.1. About this workshop report content

This document is a report on the proceedings of the workshop. The views and positions documented in this report are expressed during the workshop by participants and do not necessarily reflect IAB's views and positions.

Furthermore, the content of the report comes from presentations given by workshop participants and notes taken during the discussions, without interpretation or validation. Thus, the content of this report follows the flow and dialogue of the workshop but does not attempt to capture a consensus.

2. Workshop Scope and Discussion

The workshop was organized across three days with all-group discussion slots, one per day. The following topic areas were identified and the program committee organized paper submissions into three main themes for each of the three discussion slots. During each discussion, those papers were presented sequentially with open discussion held at the end of each day.

2.1. Session 1: Community Networks - Their Role in Internet Access of Services

The first day of the workshop focused on the role of Community Networks [RFC7962] as a way to overcome the barriers to Internet Access. Community Networks are self-organized networks wholly owned by the community and thus provide an alternative mechanism to bring connectivity and internet services to those places that lack commercial interest.

Presentations ranged from highlighting the need for measuring Quality of Experience (QoE) for Community Networks, to the potential role the Content Delivery Network (CDN) can play in Community Networks, to the role of Satellite Networks, and finally, to the vital role of the spectrum in this space.

2.1.1. The Quality of Community Networks

[MARTINEZ] highlighted the need to address Quality of Experience (QoE) in discussions around Community Networks. As a community-driven deployment, the knowledge and involvement of individuals can vary; therefore, there are no guarantees of connectivity or quality of service. There is a need to focus on user expectations and how they translate to measurable performance indicators. Further, it asks for better documenting best practices in deploying community networks as well as considering manageability considerations for community networks in protocol development. [GUIFI] as an example Community Network was discussed and some existing resources for Community Networks ([APC], [ISOC], and [TBB]) were shared by the participants.

The inconsistent quality and performance of Satellite Internet is a gap for community networks that rely on non-terrestrial networks (NTNs) for internet access [HU].

2.1.2. Strengthening Community Networks

[BENSON] focused on the prohibitive cost of the transit and Internet service for Community Networks and argued for Content Delivery Networks (CDNs) to provide transit-like and Internet services at no more than at-cost in a mutually beneficial way. Community networks still need backhaul to and from the CDNs point of presence and models for community-backhaul and open-source CDNs were highlighted. Discussion included [PANGEA] project status as well as legal and commercial considerations in such use of CDNs.

[HU] highlighted that Satellite Internet provided by advanced LEO satellite constellations can play a pivotal role in closing the connectivity gap in the urban-rural digital divide via Satellite-dependent community networks. These existing known performance and management gaps need focus to enable Satellite Internet to resolve the divide. Further, research directions such as multi-layer satellite networking, autonomous maintenance, and integration between Terrestrial and Non-Terrestrial networks were suggested.

[RENNO] called attention to the coveted 6GHz (part of the C-band with a desirable mix of coverage and capacity) as a prime choice for International Mobile Telecommunication (IMT) for 5G technology while it is in common unlicensed use in the community networks (and small ISPs). Spectrum allocations directly impact industries and market access with ramifications for community networks. Further, there was a discussion on the geopolitical tension because of it.

2.1.3. Discussion

How can the technical community address the management gap and improve best practices for Community Networks? Is the increasing complexity of the Internet making it more challenging to establish secure connections, and should this be taken into account in the design of the Internet? What steps need to be taken to make sure Community Networks are secure? Should the manageability consideration be expanded to explicitly consider Community Networks? Global Access to the Internet for All (GAIA) [GAIA] research group could be a venue for further discussion and research. Further discussion highlighted the need for readily available knowledge and tools for community networks as well as the tussle with market forces when commercial networks compete with community networks. Also, there is a lack of operational inputs from community network operators in the IETF/IRTF.

2.2. Session 2: Digital Divide - Reports and Comments

Critical internet infrastructure affects many aspects of our society significantly, although differently, the inequitable aspects of which are typically referred to as "digital inclusion" signifying that in efforts to digitalise society, there are those left out due to what is typically called the "digital divide", a related term specific to access to the Internet. These concepts together demonstrate that even if Internet connectivity is available, for some there will remain challenges towards achieving equality. This becomes especially significant as governments view the Internet as an important tool to help them reach the Sustainable Development Goals (SDG) [SDG] and to globally support human rights.

The second day of workshops was essential to understanding the nature of the digital divide. Presentations of reports interrogated at least three key aspects of the digital divide, though there is recognition that there may be more technical aspects of the digital divide that were not present. Those were: differences between population demographics in the provision of online resources by governments, inequality in the use of multilingualized domains and email addresses, and increased costs for end-user downloads of contemporary websites' sizes.

2.2.1. Disparities in service provision

Ralph Holz presented research that exposes the more limited DNS-mediated access to government websites by Indigenous communities in Australia as compared to less disadvantaged users in the same population in "Evidence for a digital divide? Measuring DNS dependencies in the context of the Indigenous population of Australia". [HOLZ] DNS dependency trends were analysed between two lists of domains serving Australian government sites for Indigenous users and the general population. Researchers found, "evidence that dependencies for the Indigenous population are indeed differently configured," indicative of a difference in service provisioning. However qualitative follow-up research is needed to interrogate both the potential reasons for these differences and whether the differences contribute to a "digital divide" that is tangible for Indigenous users.

2.2.2. Lack of consistent acceptance of language scripts

On the topic of availability of Internet services and content in multiple languages "Universal Acceptance of Domain Names and Email Addresses: A Key to Digital Inclusion" was presented by Sarmad Hussain of ICANN. [HUSSAIN] The ICANN community has increased the options for multilingual identifiers through the expansion of the Internet's DNS for use in domains and email addresses. However, while the work of technical specification and policy recommendations is complete, much work remains to deploy a multilingualized internet. Today there are around 150 internationalised domain names (IDNs) but the barriers to equal rollout of these scripts at the domain level are hindered primarily by software and applications that do not yet recognise these new scripts. "Universal Acceptance" is a programme of action for the internet community at large that can ensure IDNs are accepted and treated consistently.

2.2.3. Web Affordability and Inclusiveness

In "A Framework for Improving Web Affordability and Inclusiveness" Rumaisa Habib presented research on the connection between website size and cost to end users. [HABIB] This critical inquiry presents access in terms of affordability and through measurement demonstrates that the material costs to end users who pay for their connection based on the volume of data they download and upload have risen as the complexity of the web grows. Their research provides a framework for optimisation based on end-user affordability. This framework is anchored to reality: it proposes a fairness metric and suggests systematic adaptations to Web complexity based on "geographic variations in mobile broadband prices and income levels."

2.2.4. Discussion

These three reports discuss very different aspects of current inequalities in Internet access in various parts of the world: service provision, availability, and economic costs. Notably, the reports discuss trends that exacerbate the digital divide beyond the question of connectivity or whether users have access to the Internet, potentially bringing concrete ways that the IETF community can address digital inclusion within its remit.

Discussants noted that while there are some interesting aspects to the problem of the digital divide, such as measurements and frameworks, most of the work is getting this work to the right people at the policy layer so there is an importance of communicating this technical evidence to the right people. The IETF's role could be to build consensus on the proper solutions presented to decision-makers that put research and measurement not only in context but also in a

consensus-driven solution space. Another method to better communicate this research is by telling stories of end users in more relatable and relevant terms, which is often a challenge for the technical level and a role for more diverse stakeholders at the more local level.

2.3. Session 3: Censorship - Reports and Circumvention

This session focused on reports of censorship as observed during recent years in different parts of the world, as well as on the use of and expectation on censorship circumvention tools, mainly the use of secure VPN services.

The censorship reports, with a focus on Asia, and specifically India, as well as Russia, as an example where censorship has changed significantly recently, discussed the legal frameworks and court acts that put legal obligations on regional network providers to block traffic. Further, measurements to validate the blocking as well as analyses of how blocking is implemented were discussed, i.e. which protocols are used but also which kind of devices are used to configure the blocking rules and where are they deployed.

2.3.1. Censorship Orders, Measurements, and Device Analysis

[SAMSUDIN] reported on confirmed blocking from 10 countries (Cambodia, Hong Kong (China), India, Indonesia, Malaysia, Myanmar, Philippines, Thailand, Timor-Leste, Vietnam) in the period from 1 July 2022 to 30 June 2023. The blocking was either confirmed by OONI measurements for existing blocking fingerprints, heuristics, i.e. for new blocking fingerprints as well as news reports of blocking orders, or user experiences. Most of these countries block specific content such as porn, gambling, or certain news pages. Interestingly the blocking in Hong Kong and Myanmar is focused on the military and governmental pages of foreign countries. Blocking is often realized by either DNS tampering or HTTP tampering. For DNS, either a decided IP address, a bogon IP address (127.0.0.1), or an empty domain (nxdomain) is used. In case of DNS tampering using a decided IP address or HTTP tampering some countries provide a block page that exposes the blocking, however, more transparency about blocking is requested by civil society organizations and the iMAP project.

[GROVER] further focused the discussion on online censorship in India, Pakistan, and Indonesia. In India, where providers are responsible for implementing the blocking but no method is mandated, the six major ISPs (covering 98.82% of all subscribers) were tested on 4379 blocked websites (based on court orders, user reports, and publicly available or leaked government orders) on DNS poisoning/injection or HTTP/SNI-based censorship. Used censorship techniques

and websites blocked were different across ISPs. Multiple ISPs used two different techniques (depending on the website), and all but one provided censorship notices. Providers blocked between 1892 to 3721 (of 4379) pages with only 1115 (27.64%) of pages blocked by all ISPs. [Singh2020] In contrast, in Pakistan, the government can also order the ISPs to perform blocking and blocking has even been observed in the past on the IXP level. Since 2020, there has also been a central Web Monitoring System deployed at lines of international connectivity. In Indonesia, initially, the government guided ISPs in how to perform the blocking. The regulations were updated in 2020 to allow Indonesian ISPs to block websites at their discretion. In 2022, there was a proposal by internet service providers to centralise DNS. In Indonesia, a partial block list is publicly available, but without any indication of why something is blocked. [Grover2023]

[BASSO] reported that for Russia a high increase in additions to the Roskomnadzor's block list was observed in March 2022 as well as in December 2022, foremost covering news pages but also covering human rights organizations and social media, where more than 3500 blocking orders were added to the list by an "Unknown body". Further, blocking of domains that are not in the official Roskomnadzor's list has been observed as well.

An invited talk presented the work in [WANG] on locating censorship devices by using HTTP and TLS traceroutes, identifying device vendors through fingerprinting, and reverse-engineering censorship triggers by the use of fuzzing. E.g. for the case of Azerbaijan and Kazakhstan, they showed that a significant portion of measurements from remote countries are blocked at the endpoint, indicating local policies but connection resets are also happening in Belarus and Russia. Further, they could identify a set of commercial network devices (with filtering techniques such as firewalls) that are used in these countries for censorship and show how fuzzing can be used to fingerprint and cluster behaviours as well as potentially circumvent the deployed methods.

All speakers called for more transparency by requiring blocking messages as well as publication and auditing of blocklists. Potentially even standardization could help.

2.3.2. Use of VPNs for Censorship Circumvents and User Expectations

Further on in the session, the possibility and prevalence of using VPNs for circumvention has been discussed including user expectations and an analysis of security shortcomings of commercial VPN services. The analysis presented in [RAMESH] has shown various problems that lead to data leaks such as leakage of IPv6 traffic, non-browser traffic, or tunnel failure, not upholding user expectations, especially when used in authoritarian regimes for censorship circumvention or private communication.

The question of how common the use of VPNs for circumvention is and its legal implications, as VPNs are illegal in a few countries, has been discussed. E.g. VPNs are not officially banned in India but VPN providers need to store log data and those, who haven't complied, stopped serving India. However, more data on VPN use and blocking might be needed.

2.3.3. Discussion

After all, there is a cat-and-mouse game between censors and circumvents, however, continued work on protocol enhancements that protect user privacy is essential.

2.4. Key Take Aways

Some key takeaways from the workshop are -

- * There is a need for the technical community to address the management gaps in operating Community Networks.
- * Work should be done in documenting best practices for operating Community Networks.
- * During the development of protocols, explicit manageability considerations related to Community Networks should be considered.
- * Build consensus on solutions that have the most significant impact in fostering digital inclusion. Further, promoting these solutions ensures that efforts to bridge the digital divide are effective and inclusive.
- * Further work to enhance protocols ensuring user privacy should continue.
- * Develop further protocols (or extensions to existing protocols) that enable more transparency on filtering and promote their use and deployment.

- * Develop new VPN-like services and potentially support measurements to understand their deployment and use.
- * Further discussion of these topics could happen in GAIA, HRPC, PEARG, and MAPRG based on the relevance to the research group. The management and operations-related discussion can be taken to OPSAWG. The community could also explore if a censorship (and its circumvention) focused group could be created.

3. Informative References

- [APC] "The Association for Progressive Communications (APC)", n.d., <<https://www.apc.org/>>.
- [BASSO] Basso, S., "How Internet censorship changed in Russia during the 1st year of military conflict in Ukraine", January 2024, <<https://datatracker.ietf.org/meeting/interim-2024-biasws-03/materials/slides-interim-2024-biasws-03-sessa-online-censorship-in-india-pakistan-and-indonesia-00>>.
- [BENSON] Benson, T. A. and M. Fayed, "A \hat{A} 200\230Câ\200\231 in CDN - Access service to and from the Internet at cost for community networks", January 2024, <<https://www.ietf.org/slides/slides-biasws-a-c-in-cdn-access-service-to-and-from-the-internet-at-cost-for-community-networks-00.pdf>>.
- [GAIA] "Global Access to the Internet for All Research Group", n.d., <<https://www.irtf.org/gaia.html>>.
- [GROVER] Grover, G., "Online censorship in India, Pakistan and Indonesia", January 2024, <<https://datatracker.ietf.org/meeting/interim-2024-biasws-03/materials/slides-interim-2024-biasws-03-sessa-online-censorship-in-india-pakistan-and-indonesia-00>>.
- [Grover2023] Grover, G. and C. Cath, "The infrastructure of censorship in Asia", October 2023, <<https://archive.org/details/eaten-by-the-internet/>>.
- [GUIFI] "Guifi.net", n.d., <<https://guifi.net/en>>.
- [HABIB] Habib, R., Tanveer, S., Inam, A., Ahmed, H., and A. Ali, "A Framework for Improving Web Affordability and Inclusiveness", September 2023, <<https://www.ietf.org/slides/slides-biasws-a-framework-for-improving-web-affordability-and-inclusiveness-00.pdf>>.

- [HOLZ] Holz, R., Nazemi, N., Tavallaie, O., and A. Y. Zomaya, "Evidence for a digital divide? Measuring DNS dependencies in the context of the indigenous population of Australia", 2023, <<https://www.ietf.org/slides/slides-biasws-evidence-for-a-digital-divide-measuring-dns-dependencies-in-the-context-of-the-indigenous-population-of-australia-00.pdf>>.
- [HU] Hu, P., "Closing the Performance and Management Gaps with Satellite Internet - Challenges, Approaches, and Future Directions", January 2024, <<https://www.ietf.org/slides/slides-biasws-closing-the-performance-and-management-gaps-with-satellite-internet-challenges-approaches-and-future-directions-01.pdf>>.
- [HUSSAIN] Hussain, S., "Universal Acceptance of Domain Names and Email Addresses - A Key to Digital Inclusion", 2023, <<https://www.ietf.org/slides/slides-biasws-universal-acceptance-of-domain-names-and-email-addresses-a-key-to-digital-inclusion-01.pdf>>.
- [ISOC] "Community networks help bridge the connectivity gap", n.d., <<https://www.internetsociety.org/action-plan/community-networks/>>.
- [MARTINEZ] MartÃ-nez-Cervantes, L. M. and R. Guevara-MartÃ-nez, "Community Networks and the Quest for Quality", January 2024, <<https://www.ietf.org/slides/slides-biasws-community-networks-and-the-quest-for-quality-00.pdf>>.
- [PANGAEA] "Project Pangea from Cloudflare", n.d., <<https://www.cloudflare.com/en-gb/pangea/>>.
- [RAMESH] Ramesh, R., "Investigating the VPN Ecosystem through the lens of Security, Privacy, and Usability", January 2024, <<https://datatracker.ietf.org/meeting/interim-2024-biasws-03/materials/slides-interim-2024-biasws-03-sessa-investigating-the-vpn-ecosystem-through-the-lens-of-security-privacy-and-usability-00>>.
- [RENNO] RennÃ³, R., "Maximising Connectivity - The Spectrum's Vital Role in Technology Access", January 2024, <<https://www.ietf.org/slides/slides-biasws-position-paper-by-raquel-renno-01.pdf>>.

- [RFC7962] Saldana, J., Ed., Arcia-Moret, A., Braem, B., Pietrosevoli, E., Sathiaselan, A., and M. Zennaro, "Alternative Network Deployments: Taxonomy, Characterization, Technologies, and Architectures", RFC 7962, DOI 10.17487/RFC7962, August 2016, <<https://www.rfc-editor.org/rfc/rfc7962>>.
- [SAMSUDIN] Samsudin, S., "iMAP (Internet Monitoring Action Project) 2023 Internet Censorship Report", January 2024, <<https://www.ietf.org/slides/slides-biasws-position-paper-by-raquel-renno-01.pdf>>.
- [SDG] "Sustainable Development Goals", n.d., <<https://sdgs.un.org/goals>>.
- [Singh2020] Singh, K., Grover, G., and V. Bansal, "How India Censors the Web", July 2020, <<https://dl.acm.org/doi/abs/10.1145/3394231.3397891>>.
- [TBB] "Tribal Broadband Bootcamp", n.d., <<https://tribalbroadbandbootcamp.org/>>.
- [WANG] Raman, R. S., Wang, M., Dalek, J., Mayer, J., and R. Ensafi, "Network Measurement Methods for Locating and Examining Censorship Devices", November 2023, <<https://datatracker.ietf.org/meeting/interim-2024-biasws-03/materials/slides-interim-2024-biasws-03-sessa-online-censorship-in-india-pakistan-and-indonesia-00>>.

Appendix A. Position Papers

19 position papers were submitted to the workshop call for papers. 11 were selected for publication. Papers that were not published either (1) only provided a very prelimited analysis of an idea that was felt to be incomprehensive for discussion at the workshop, or (2) addressed problems that were beyond the scope as dedicated for the workshop discussion e.g. discussing cyber security threads as a barrier for participation or implication of technology in regulation that imposes blocking. Both of these topics pose a potentially severe risk on the open Internet, however, these risks might provide a high risk for all Internet users but do not necessarily imply an unbalance.

All accepted papers are available at:
<https://datatracker.ietf.org/group/biasws/materials/>

This is the list of all published papers:

Community Networks:

- * L. M. MartÃ-nez-Cervantes, R. Guevara-MartÃ-nez: Community Networks and the Quest for Quality [MARTINEZ]
- * T. Benson, M. Fayed: A â\200\230Câ\200\231 in CDN: Access service to and from the Internet for community networks at-cost [BENSON]
- * P. Hu: Closing the Performance and Management Gaps with Satellite Internet: Challenges, Approaches, and Future Directions [HU]
- * R. RennÃ³: Maximising Connectivity: The Spectrum's Vital Role in Technology Access [RENNO]

Digital Divide:

- * R. Holz, N. Nazemi, O. Tavallaie, A.Y. Zomaya: Evidence for a digital divide? Measuring DNS dependencies in the context of the indigenous population of Australia [HOLZ]
- * S. Hussain: Universal Acceptance of Domain Names and Email Addresses: A Key to Digital Inclusion [HUSSAIN]
- * R. Habib, S. Tanveer, A. Inam, H. Ahmed, A. Ali, Z.A. Uzmi, Z.A. Qazi, I.A. Qazi: A Framework for Improving Web Affordability and Inclusiveness [HABIB]
- * J. Ott, G. Bartolomeo, M.M. Bese, R. Bose, M. Bosk, D. Guzman, L. KÃ¼rkkÃ¼inen, M. Kosek, N. Mohan: The Internet: Only for the Fast (and Furious)?
- * L.Y. Ohlsen: BIAS workshop - M-Lab Position Paper submission

Censorship:

- * S. Nurliza Samsudin: iMAP (Internet Monitoring Action Project) 2023 Internet Censorship Report [SAMSUDIN]
- * G. Grover: The infrastructure of censorship in Asia [Grover2023]
- * S. Basso: How Internet censorship changed in Russia during the 1st year of military conflict in Ukraine [BASSO]

In addition to the submitted paper two invited talks were presented based on published papers:

- * R. Sundara Raman, M. Wang, J. Dalek, J. Mayer, R. Ensafi:
Network Measurement Methods for Locating and Examining Censorship
Devices [WANG]
- * R. Ramesh, A. Vyas, R. Ensafi: [\200\234All of them claim to be the
best\200\235: A multi-perspective study of VPN users and VPN providers](#)

Appendix B. Workshop Participants

The workshop participants were Arnaud Taddei, Carlos Pignataro, Carsten Bormann, Cindy Morgan, Colin Perkins, Cory Myers, Dan Sexton, David Guzman, David Millman, David Schinazi, Dhruv Dhody, Gurshabad Grover, Hanna Kreitem, Jane Coffin, Jiankang Yao, Jörg Ott, Juan Peirano, Lai Yi Ohlsen, Luis Martinez, Mallory Knodel, Marwan Fayed, Matthew Bocci, Michael Welzl, Michuki Mwangi, Mirja Kuehlewind, Mona Wang, Peng Hu, Ralph Holz, Raquel Renno, Reethika Ramesh, Rumaisa Habib, Sarmad Hussain, Simone Basso, Siti Nurliza Samsudin, Suresh Krishnan, Theophilus Benson, Tirumaleswar Reddy, Tommy Pauly, Vesna Manojlovic, and Wes Hardaker.

Appendix C. Workshop Program Committee

The workshop program committee members were Christopher Wood (IAB, Cloudflare), Dhruv Dhody (IAB, Huawei), Mallory Knodel (IAB, Center for Democracy and Technology), Mirja Kuehlewind (IAB, Ericsson), and Tommy Pauly (IAB, Apple).

IAB Members at the Time of Approval

Internet Architecture Board members at the time this document was approved for publication were: TODO

Acknowledgments

Thanks to Arnaud Taddei for helpful suggestions to improve this report.

Authors' Addresses

Mirja Kuehlewind
Email: ietf@kuehlewind.net

Dhruv Dhody
Email: dd@dhruvdhody.com

Mallory Knodel

Email: mknodel@cdt.org