

# Measurement and Analysis for Protocols

Research Group (maprg)

## Special Session on AI Crawler Traffic Impacts

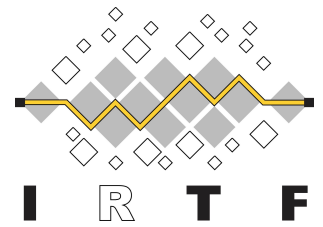
July 25, 2025, IETF123 in Madrid

co-chairs <maprg-chairs@ietf.org>:

Mirja Kühlewind <[mirja.kuehlewind@ericsson.com](mailto:mirja.kuehlewind@ericsson.com)>

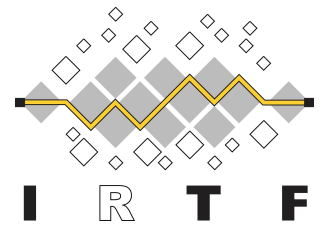
Dave Plonka <[plonka@wiscnet.net](mailto:plonka@wiscnet.net)>

# Note Well – Privacy & Code of Conduct



- As a participant in, or attendee to, any IRTF activity you acknowledge that written, audio, video, and photographic records of meetings may be made public
- Personal information that you provide to IRTF will be handled in accordance with the Privacy Policy at <https://www.ietf.org/privacy-policy/>
- As a participant or attendee, you agree to work respectfully with other participants; please contact the ombudsteam (<https://www.ietf.org/contact/ombudsteam/>) if you have questions or concerns about this
- See [RFC 7154](#) (Code of Conduct) and [RFC 7776](#) (Anti-Harassment Procedures), which also apply to IRTF
- Also see [RFC 9775](#) (IRTF Code of Conduct)

# Goals of the IRTF



- The Internet Research Task Force (IRTF) focuses on longer term research issues related to the Internet while the parallel organisation, the IETF, focuses on shorter term issues of engineering and standards making
- **The IRTF conducts research; it is not a standards development organisation**
- While the IRTF can publish informational or experimental documents in the RFC series, its primary goal is to promote development of research collaboration and teamwork in exploring research issues related to Internet protocols, applications, architecture, and technology
- See “An IRTF Primer for IETF Participants” – [RFC 7418](#)

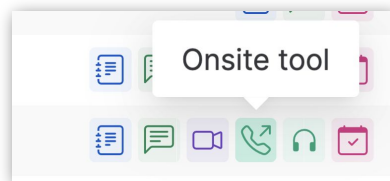
# Administrativa

- Charter: <https://datatracker.ietf.org/group/maprg/charter/>
- Mailing List: [maprg@irtf.org](mailto:maprg@irtf.org)  
Subscriptions: <https://www.irtf.org/mailman/listinfo/maprg>
- Today's slides:  
<https://datatracker.ietf.org/meeting/123/session/maprg/>
- Meetecho: <https://meetecho.ietf.org/client/?group=maprg>
  - Please join Meetecho (lite or full version) from the room to sign in for the bluesheets (and enter the queue)!

# Meeting Tips

## In-person participants

- Make sure to sign into the session via Datatracker or the QR Code in this session.
- Use Meetecho (usually the "Meetecho lite") client to:
  - join the mic queue
  - participate in shows of hands
- *Keep audio and video off if not using the onsite version.*



## Remote participants

- Make sure your audio and video are off unless you are chairing or presenting during a session.
- Use of a headset is strongly recommended.

**This session is being recorded**

# Note about this session

- MAPRG has two 1.5h hour session at IETF-123
  - Our regular session is on Friday 9:30-11:00!
- Measuring crawler traffic is not directly measuring protocols (our scope) but there is ongoing work on crawler behavior in the IETF
  - The AIPref group was chartered earlier this year to define a vocabulary for preferences of content usage for e.g. AI training
  - This information can the be added to the Robots Exclusion Protocol (RFC9309) aka robots.txt or other containers
- Beside content usage concern, there is also a concern about increased crawler load, especially on open sources-related pages
  - This session will assess the current state and consider mitigation approaches
  - Note the Web Bots Auth BoF on Monday is also related to this problem.

# Other related things

- Crawler best practices: [draft-illyes-aipref-cbcp-00](#)
- IndexNow <https://www.indexnow.org/> (later on the agenda)
- Blog posts
  - [How crawlers impact the operations of the Wikimedia projects](#)
  - [Please stop externalizing your costs directly into my face](#)
  - [Trapping misbehaving bots in an AI Labyrinth](#)
  - [No hallucinations here: track the latest AI trends with expanded insights on Cloudflare Radar](#)
- New article
  - [FOSS infrastructure is under attack by AI companies](#)
  - [Open source devs say AI crawlers dominate traffic, forcing blocks on entire countries](#)

# Agenda

- 9:30** Overview and Status - *Mirja/Dave*
- 9:35** Web Crawl Refusals: Insights From Common Crawl - Mostafa Ansar (remote)
- 9:45** Somesite I Used To Crawl: Awareness, Agency and Efficacy in Protecting Content Creators From AI Crawlers - Elisa Luo (remote)
- 10:00** Automated traffic affects IETF services - Robert Sparks
- 10:05** Bot Traffic at Wikimedia: Measurement, Identification, and Response  
- Chris Petrillo and Birgit Müller (remote)
- 10:25** AI Crawlers - Insights from Cloudflare - Thibault Meunier
- 10:30** IndexNow: A Real-Time Protocol for Measurable Web Indexing Efficiency - Krishna Madhavan