

Taking stock:
questions and hacking groups

IAB Workshop on Analysing IETF Data (AID)

Chair: Mirja Kühlewind - Nov 29, 2021

Tools, data, and methods

Data sources

- List achieves: email content and subscription data
- Datatracker: Draft/RFC content and metadata (including action history); user account information; meeting proceedings and minutes; wg metadata
- ietf.org: meeting participation

Tools, data, and methods

Tools

- **AUTHORSTATS** (<https://www.arkko.com/tools/allstats/>) and **IETF document statistics** (<https://www.arkko.com/tools/docstats.html>): datatracker interfaces
- **The ietfdata Library**: datatracker, RFC index, email achieve
- **The RFC Prolog Database**: RFC content and metadata as well as outcomes of tools integrated in the datatracker or IETF process
- **BigBang project**: email subscriptions and user account data from datatracker

Tools, data, and methods

Challenges to improve tooling (for authors)

- Authors are interested in how their publications are referenced or about getting an easy listing of what documents they have, companies are interested in how their efforts and topics compare to what other companies are doing, for diversity it is interesting to understand geographical split of various levels of activities, same for gender distribution, and so on. [Arkko]
- What are other relevant information to display together with an RFC? [Hoffman]

Participation trends

Diversity and Inclusivity

- How to identify the International footprint of communication within the IETF? How to identify temporal diversity characteristics in international participation (by analysing participation trends within IETF communication over time)? [Hardaker]
- Would feminists have built a better internet? Better understand 1) Diversity of participation in internet governance, 2) Influence and marginalization in internet governance, 3) Alignment between expressed governance values and feminist principles [Knode]
- Have IETF diversity initiatives resulted in greater diversity among IETF standards authorship? Internet Standards Truly Diverse? Which firms and industry sub-sectors have the strongest influence on IETF standards development? [Le]
- How to derive actionable insights to facilitate greater diversity, inclusion and fairness in its operations? [Sinha]
- How findings help to create a more transparent and inclusive standard development process? [Yachmeneva]
- How to improve gender diversity in the selection process? [Tarantino]
- Agree on a shared definition of active IETF participation, and then attempt to identify classes of participants that share common participation patterns. [Eggert]

Participation trends

Affiliation and Industry Control

- How to better understand global competition in leadership position? [Baron]
- Which internet standard organisations are formally-institutionally and actually autonomous from corporate interests? [Antoine]
- Whether it is possible to detect, empirically, when an organization is acting within the IETF through individuals that it employs, as opposed to letting those employees act in a more personal capacity? [Benthall]
- Are there other data sets or existing tools we can use for understanding and analyzing the organizations that mediate participation in governance fora? [Doty]
 - Data on companies and their financial status, location, industry sector, etc. could all be useful supplements.
 - It might be useful to have a simple common format for information on participants, including affiliations at points of time, so we can share findings from across different standard-setting or other Internet governance venues.

Publications, process, and decision-making

Methods used for content analysis

- Time series analysis of keywords (used in emails) [Andersdotter]
- Coreference resolution (CR) of (IETF) mailing list content [Groth]
- RFC content and metadata analysis/correlation [McFadden]
- Volume analysis over time (email subscription and author data in RFCs) [Castro]
- Correlation of discussion with the development of drafts (email content) [Griwodz]
- Natural Language Processing (NLP) techniques (of email content) [Welzl]

Publications, process, and decision-making

Question/Challenges for (Email and RFC) content analysis

- Which other keywords could/should be applied for time series analysis of mailing list traffic? How to correlate this data to events? [Andersdotter]
- What are the challenges faced in analyzing IETF data using new natural language processing approaches? [Groth]
- How to use these analyses to improvise (security) work in the IETF?
 - How and why security vulnerabilities manifest themselves from design to implementation? How and where security improvements could be made to the RFC process? How we can utilise the IETF data for the baselining and improvement of security requirement specification within the RFC process? [McFadden]

Publications, process, and decision-making

IETF process: Decision making and success factor

- Understanding Internet Protocol Design Decisions — Better understand _why_ it works the way it does? [Welzl]
- Which factors lead to successful uptake and deployment of protocols, and derive insights to improve the standardisation process? [Castro]
- What are the reason for a failure (or success) of a proposed draft? [Griwodz]
 - Does seniority improve the chances of success? But do they face less or more resistance in discussions, or is the atmosphere in discussions different from that faced by newcomers? Does the atmosphere in discussions for proposals that face active resistance develop over time, or does it stagnate?

IETF Organization

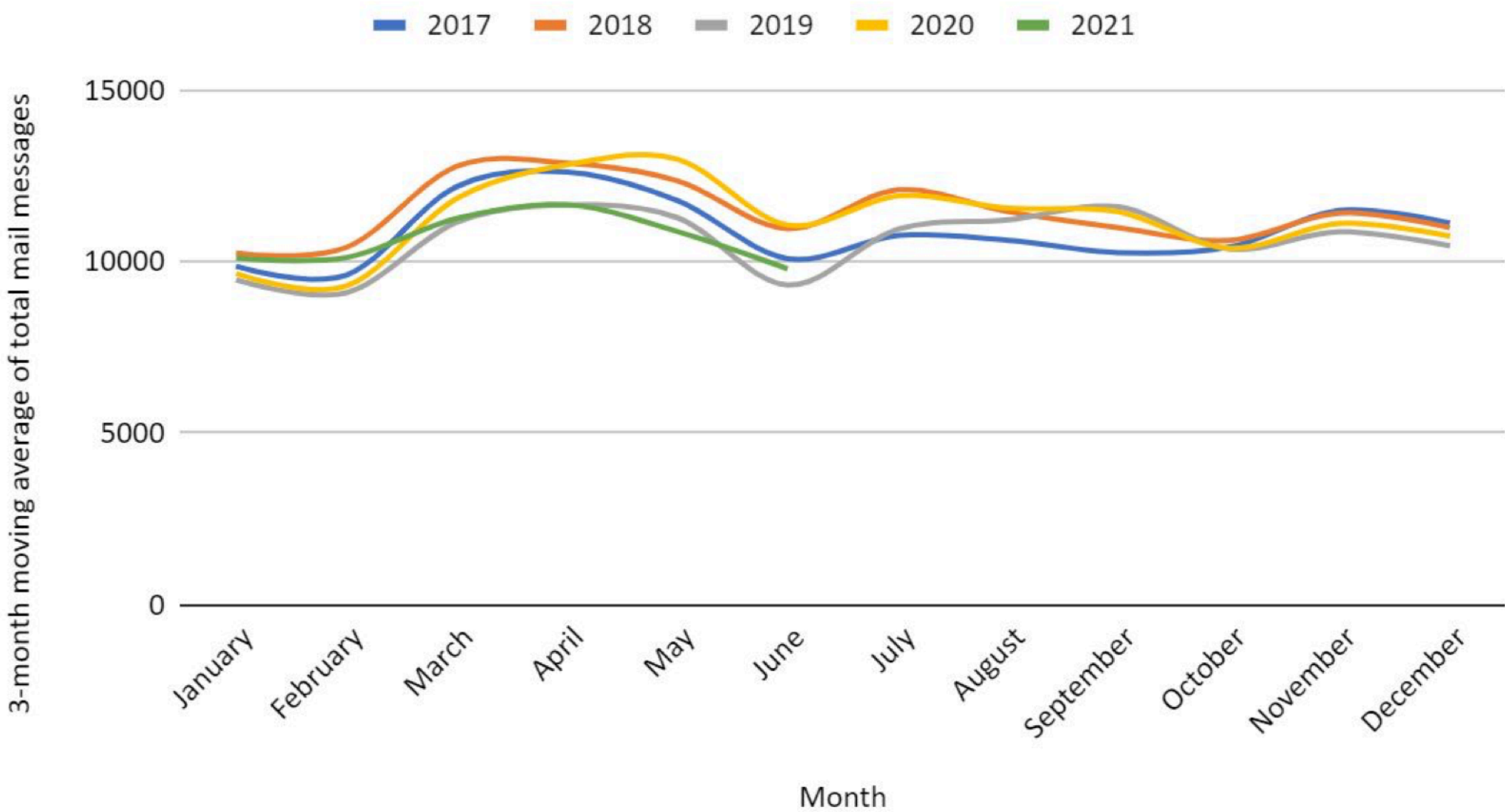
Sustainability and Climate change

- What the IETF's role ought to be, can be, and is with respect to climate change? [Becker]
- How does air traffic CO₂eq emission of IETF meetings evolve? Better understand how to make the IETF grow (to define a meeting strategy) [Migault]

IETF Meetings and COVID-19

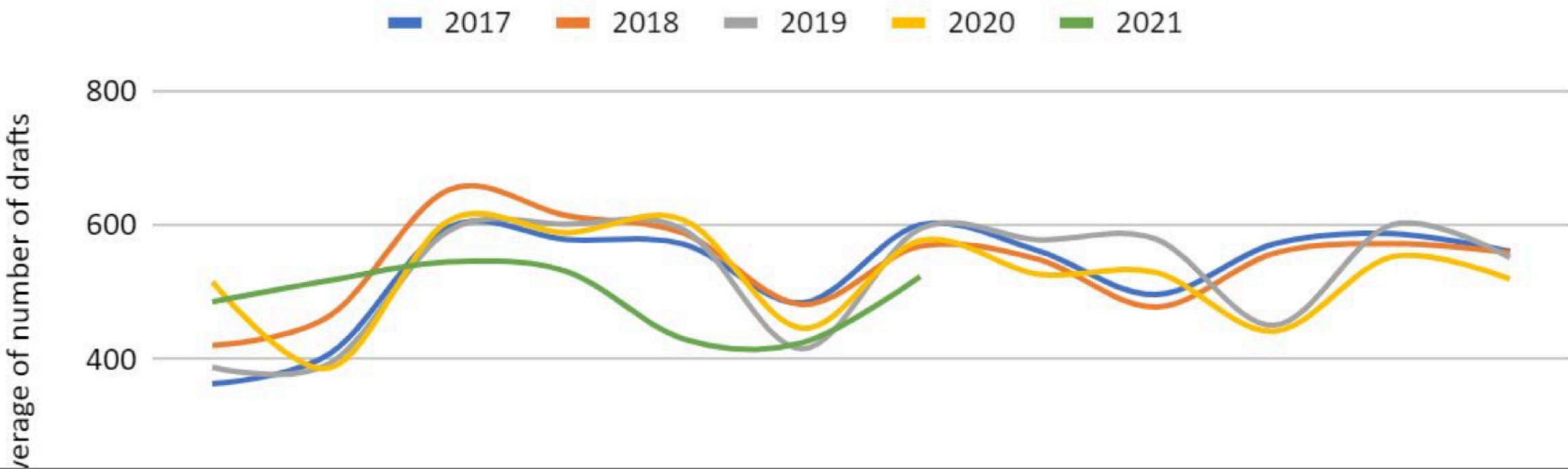
Data presented at IETF-111

Mail messages

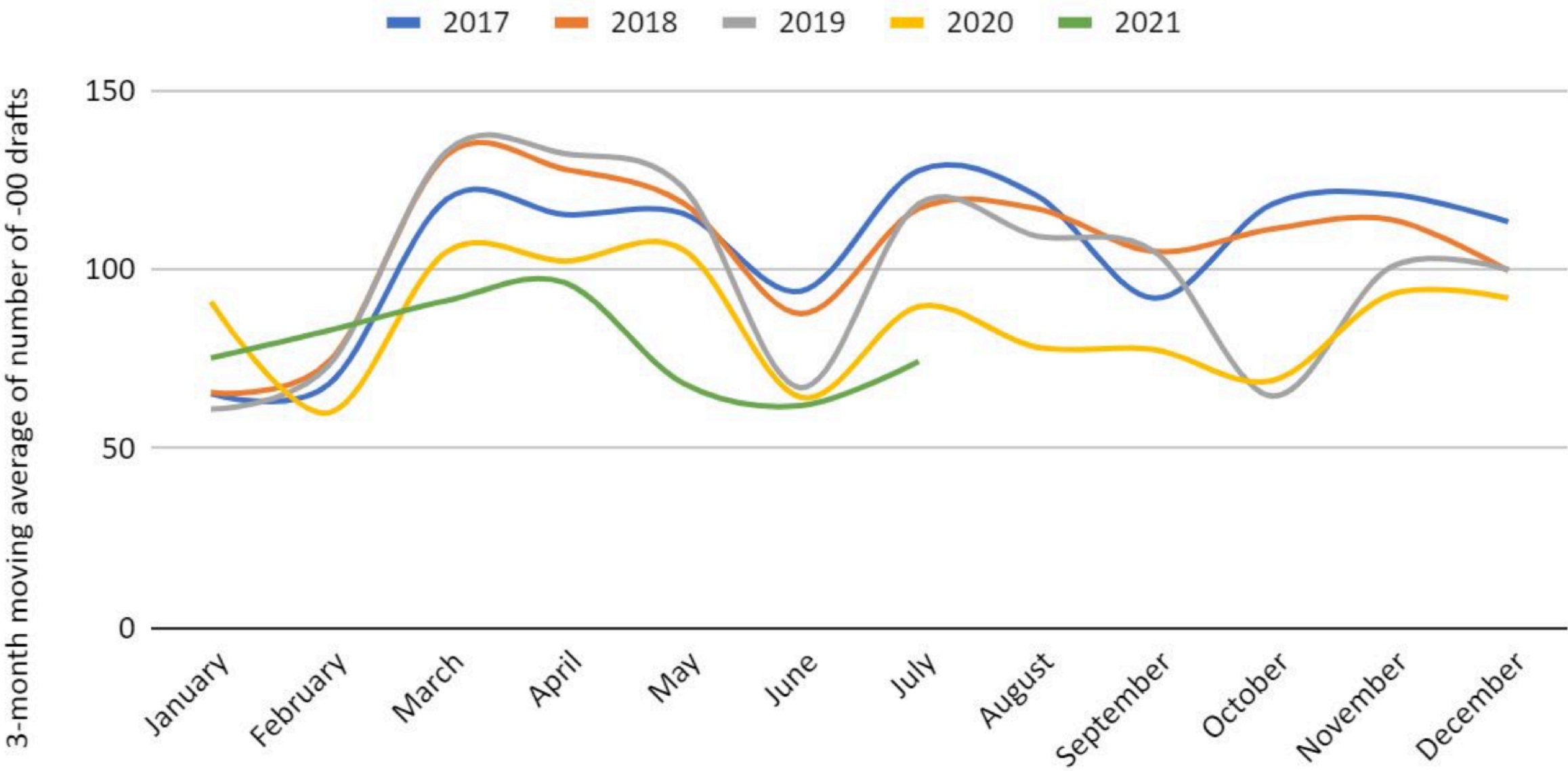


Note: excludes "quic-issue"

Internet-drafts published



-00 Internet-drafts



IETF Meetings and COVID-19

Additional questions from draft-kuehlewind-shmoo-online-meeting

- What impact had the move to fully online meeting on productivity in general but particularly new work?
- How did use of interim meetings change since the start of the pandemic? How does that impact productivity/draft submissions? How does that impact cross-area interactions?
- How did meeting participation change (with the move to fully online meetings)? Did the number of newcomers change? Has diversity increased? Do regular participants participate more/less often? How many sessions do participants visit? Which sessions are often visited in parallel by the same participant? Are more/less people actively speaking up at the mic line? Are newcomers actively participating in the session?
- How is gather used? Do people talk only to know people? However have people group conversation vs. one-on-one chats?