

Characterizing the IETF through its consensus mechanisms

PRIYANKA SINHA, Tata Consultancy Services Limited, Indian Institute of Technology Kharagpur, India

MICHAEL ACKERMANN, Industry Network Technology Council, USA

PABITRA MITRA, Indian Institute of Technology Kharagpur, India

ARVIND SINGH, AI & Robotics Technology Park, India

AMIT KUMAR AGRAWAL, Tata Consultancy Services Limited, India

We propose analysis into the consensus mechanism of IETF [7] that promotes internet drafts into RFCs. Using the data publicly available from IETF, such as mailing lists, internet draft action history, minutes of meeting, affiliation records, etc, we identify various mechanisms with which to characterize its dynamics. Through the use of novel text mining, time series clustering, graph mining and psycholinguistic approaches to understand the consensus mechanism within IETF deeply, we propose to derive actionable insights to facilitate greater diversity, inclusion and fairness in its operations.

Additional Key Words and Phrases: ietf, consensus, fairness, diversity, inclusion, social network analysis, text mining, psycholinguistics, time series clustering

1 INTRODUCTION

IETF is an open, inclusive platform for developing internet standards and their uptake in practice. It has a large participation of key technical persons keeping the internet running. Primary participation in the IETF is through writing of IETF drafts to propose

Authors' addresses: Priyanka Sinha, priyanka.sinha.iitg@gmail.com, Tata Consultancy Services Limited, and Indian Institute of Technology Kharagpur, Kolkata, India; Michael Ackermann, Industry Network Technology Council, USA, mackermann@bcbsm.com; Pabitra Mitra, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal, India, pabitra@gmail.com; Arvind Singh, arvind1.singh@gmail.com, AI & Robotics Technology Park, India; Amit Kumar Agrawal, amitk.agrawal@tcs.com, Tata Consultancy Services Limited, India.

new protocols or modifications to previous ones as the internet and its use evolves. Once a draft is proposed, it goes through a well established process [2] to gain consensus in becoming an RFC and eventually an internet standard.

The IETF facilitates these processes through mailing lists, working groups, physical meetings and conferences. They have wide participation from industry as well as academia around the world.

We propose to study the following research problems in characterizing the IETF through our efforts:

- Whether there is increased diversity, inclusion and wider participation due to virtual meetings versus physical meetings, especially for the consensus mechanisms. This would help generate insights for the IETF to improve engagement.
- Mine the organizational structure of the IETF from communication such as mailing lists, meeting minutes, internet draft action history, affiliations, etc. This would help regularly uncover homophilic groups, leaders and influence mechanisms to enable persons relatively new to the IETF to discover advocates thereby improving overall efficiency.

The [1] study demonstrates that there is low diversity in the participation along various demographic parameters, exhibiting long tailed behavior. 7% of IETF RFC authors are affiliated to Cisco, more than 20% of RFC authors are from USA based commercial networking equipment manufacturing enterprises, while almost all internet drafts are proposed from authors either in USA, Europe or China.

Based on the analysis by [3] and [5, 6], it is evident that senior members are more active on the mailing lists and therefore may have a greater influence. It is expected then, that dialogue resulting in consensus for a newbie to the IETF would require advocacy from senior members irrespective of technical merit. There are well known long standing contentious issues that are debated regularly. It is not yet known if the delay in consensus is due to factions or lack of technical clarity. It is not also known, whether there are groups of members who are regularly opposed and consistently fail to obtain consensus despite technical merit.

2 IMPACT OF PHYSICAL MEETINGS ON CONSENSUS

Based on our preliminary observation of consensus, we observe that initial consensus and advocacy for the draft is achieved outside of the mailing lists, perhaps by word of mouth or in meetings. The need for physical meetings for endorsements can be a serious deterrent to inclusiveness and fairness in the consensus process. These meetings have been physical meetings before Covid19, whose minutes are recorded and available openly. This advocacy allows favorable response and higher chance of consensus into an RFC. With the Covid19 pandemic upon us, the IETF has held many virtual meetings for its various working groups as well as conferences. By studying the volume, diversity and consistency of engagement on mailing lists and meeting minutes historically versus recently, we can understand whether the virtual meetings has improved the diversity and inclusion of participation while lowering the impedance to seeking consensus. This study would allow us to generate actionable insights for the IETF.

3 CHARACTERIZING THE STRUCTURE OF THE IETF

The IETF does not have any defined hierarchy or labels governing the consensus mechanism or the advocacy. Based on our preliminary observations, we propose to characterize and mine, irrespective of demographic parameters, the organizational structure within the IETF, as it exists today, that influences the consensus mechanism itself as well as study its evolution.

We would obtain historical data from IETF using the [8] tool and rsync [12] such as mailing lists, meeting minutes, internet draft action history, author affiliations [1]. For our study we already are aware of well known long standing contentious issues that are debated regularly. For example, those surrounding IPv6 such as hop by hop extension headers and /64 interface prefix to /56, especially on the v6ops mailing list. The co-authorship, affiliations and communication on the mailing list and in meetings can be represented as a multi layered graph that can be mined using novel multi-edged node embeddings to mine groups taking into account psycholinguistic [4] and sentiment [9] attributes as well. This would provide us a data driven view of the structure within IETF as it exists and its evolution. Being a computational approach, we can execute this analysis at regular intervals to observe the evolution of the structure as well and update accordingly.

By mining the personality traits [10] of members unobtrusively from their communication, we can create actionable insights to the IETF on nudges for greater inclusiveness. By clustering the members on the time series of their engagement in the IETF [11], we shall be able to identify homophilic groups and their consistency. This would provide essential information to the IETF members on whom to reach out to as birds of a feather (BoF).

From an academic standpoint, the IETF is an open efficient enterprise and thus allows the pursuit of developing universal enterprise behavior models mined out of the dialogues within it that can be of use widely.

ACKNOWLEDGMENTS

The authors thank Nalini Elkins, Dhruv Dhody and Mohit P. Tahiliani for their encouragement as well as their insightful and helpful reviews.

REFERENCES

- [1] Jari Arkko. [n. d.]. Document Statistics : What’s Going on in the IETF? <https://www.arkko.com/tools/docstats.html>
- [2] Brian Carpenter. [n. d.]. The IETF process: an informal guide. <https://www.ietf.org/standards/process/informal/>
- [3] Ignacio Castro, Colin Perkins, Gareth Tyson, Matthew Purver, Patrick Healey, Mladen Karan, Prashant Khare, Stephen McQuistin, and Waleed Iqbal. [n. d.]. 50 Years of Standardisation in the Internet. https://csperkins.org/research/protocol-standards/2020-12-10-ignacio-iesg-talk/2020-12-10_IESG-50-years-IETF-send.pdf
- [4] Ethan Fast, Binbin Chen, and Michael S. Bernstein. 2016. Empath: Understanding Topic Signals in Large-Scale Text. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI '16*). Association for Computing Machinery, New York, NY, USA, 4647–4657. <https://doi.org/10.1145/2858036.2858535>
- [5] Prashant Khare. [n. d.]. Impact of early engagement on longevity of IETF participation. <https://sodestream.github.io/impact-of-early-engagement-on-longevity-of-ietf-participation.html>
- [6] Stephen McQuistin, Mladen Karan, Prashant Khare, Colin Perkins, Gareth Tyson, Matthew Purver, Patrick Healey, Waleed Iqbal, J. Qadir, and Ignacio Castro. 2021. Characterising the IETF Through the Lens of RFC Deployment. In *Proceedings of the ACM Internet Measurement Conference (IMC '21)*. Association for Computing Machinery, New York, NY, USA.
- [7] Pete Resnick. 2014. On Consensus and Humming in the IETF. RFC 7282. <https://doi.org/10.17487/RFC7282>

- [8] Nick Doty Sebastian Benthall, Niels ten Oever. [n. d.]. BigBang : Analyzing open source communities. <http://dataactive.github.io/bigbang/>
- [9] Priyanka Sinha, Anirban Dutta Choudhury, and Amit Kumar Agrawal. 2014. Sentiment Analysis of Wimbledon Tweets. In *4th Workshop on Making Sense of Microposts (#Microposts2014) collocated with WWW 2014*. 51–52. http://ceur-ws.org/Vol-1141/paper_10.pdf
- [10] Priyanka Sinha, Lipika Dey, Pabitra Mitra, and Anupam Basu. 2015. Mining HEXACO personality traits from Enterprise Social Media. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Association for Computational Linguistics, Lisboa, Portugal, 140–147. <http://aclweb.org/anthology/W15-2920>
- [11] Priyanka Sinha, Lipika Dey, Pabitra Mitra, and Dilys Thomas. 2020. A Hierarchical Clustering Algorithm for Characterizing Social Media Users. In *Companion Proceedings of the Web Conference 2020 (Taipei, Taiwan) (WWW '20)*. Association for Computing Machinery, New York, NY, USA, 353–362. <https://doi.org/10.1145/3366424.3383296>
- [12] Andrew Tridgell, Paul Mackerras, and Wayne Davison. 2021. Rsync. https://en.wikipedia.org/wiki/Rsync#cite_note-23