# Poster: Characterising the IETF Through the Lens of RFC Deployment


Stephen McQuistin
sm@smcquistin.uk
University of Glasgow

Mladen Karan
m.karan@qmul.ac.uk
Queen Mary University of London

Prashant Khare
p.khare@qmul.ac.uk
Queen Mary University of London

Colin Perkins
csp@csperkins.org
University of Glasgow

Gareth Tyson
g.tyson@qmul.ac.uk
Queen Mary University of London

Matthew Purver
m.purver@qmul.ac.uk
Queen Mary University of London

Patrick Healey
p.healey@qmul.ac.uk
Queen Mary University of London

Waleed Iqbal
w.iqbal@qmul.ac.uk
Queen Mary University of London

Junaid Qadir
junaid.qadir@itu.edu.pk
ITU Lahore

Ignacio Castro
i.castro@qmul.ac.uk
Queen Mary University of London



## ABSTRACT

Protocol standards, defined by the Internet Engineering Task Force (IETF), are crucial to the successful operation of the Internet. In this poster, we'll present some of the important highlights of our paper, "*Characterising the IETF Through the Lens of RFC Deployment*" [3], in which we conducted a large-scale empirical study of IETF activities. Using a unique dataset, we examine the shifts and trends within the standards development process, showing how protocol complexity and time to produce standards has increased over time. We develop statistical models of factors that lead to successful uptake and deployment of protocols.


## 1 TRENDS IN THE IETF

The IETF is an open standards organisation, which develops Internet standards via contributions and collaborations across a number of voluntary stakeholders. Most of the IETF's day-to-day work is conducted on public mailing lists, in conjunction with three plenary meetings, and numerous interim working group meetings per year.

We make use of data from a number of sources, including the Datatracker, RFC Editor, and IETF IMAP mail archive. Our paper [3] describes these sources, and the tooling that we've made available for accessing them. We examine metadata for up to 5,707 RFCs. We analysed 2,439,240 messages, sent from 74,646 unique email addresses, across 1,153 lists. We observe a number of important trends:

***Declining RFC production.*** As shown in Figure 1, there has been a decline in the output of the IETF, in terms of RFCs, in recent years, having peaked in 2005. While output has been declining, the time taken to publish an RFC has grown. Figure 2 plots the median number of days from the

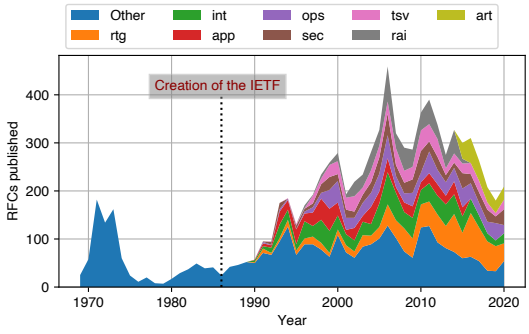



**Figure 1: RFCs by area. *Other* includes legacy RFCs, and RFCs from other non-IETF streams.**

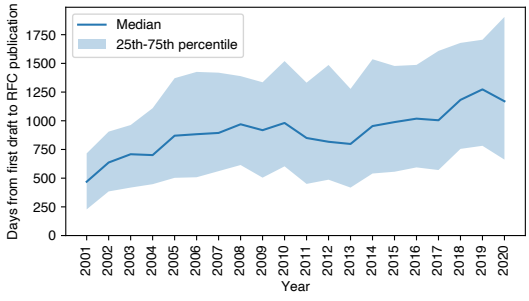



**Figure 2: Days from first draft to RFC publication**

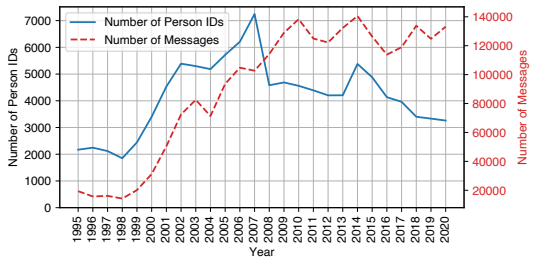



**Figure 3: Trends in mail volume and participants**

submission of an RFC's first draft, through to its publication as an RFC. This shows a clear trend: the median number of days to publication was 469 in 2001, rising to 1,170 in 2020.

***Geodistribution of authors.*** We found that North America, while still over represented, is becoming less dominant. 75% of authors were from North America in 2001, and this has declined to 44% in 2020. Representation of both Europe and Asia has grown, from 17% to 40% and 6% to 14% respectively.

***Author affiliations.*** Cisco remains a consistent employer of IETF contributors, with around 12% of authors affiliated with the company in 2020, and having been the single largest affiliation across all years in the dataset. We can also see the rise of Huawei beginning in 2005, with 7.1% of authors being affiliated with the company in 2020, having peaked at 9.7% in 2018. Google has a similar trajectory, first appearing in the dataset in 2006, with 3.8% of authors in 2020.

***Volume of emails.*** Figure 3 presents the number of emails sent across the last 25 years, showing that e-mail volumes have grown significantly with time, plateauing at around 130,000 annual messages since around 2010. The figure also shows that the number of unique people observed every year in the mail lists is in decline.

## 2 PREDICTING DEPLOYMENT

The trends that we observed highlight that, over time, the complexity of the standardisation process is increasing: RFCs take longer to produce, and relate to a larger number of other documents; authors come from a changing and diversifying set of affiliations and countries; and the volume of email interaction for each RFC is increasing. We next seek to understand how these factors might impact the successful deployment of an RFC in-the-wild, and to test if such success can be predicted. To do this, we developed a statistical model.

From our modelling, we found that building on existing RFCs, having a limited scope, and having well-specified requirements positively correlated with deployment. Our results broadly support the conclusions of [6], where the IAB noted that, in addition to being technically sound, a successful protocol must meet a need, be incrementally deployable, and be open to extension and maintenance.

## 3 RELATED WORK

Jari Arkko maintains a website[1] and tooling that provides various statistics about the IETF. BigBang [1] is a Python toolkit for analysing mailing list data. Niedermayer et al. [4] discuss the challenges of working with large mailing list datasets. Huitema [2] carried out an evaluation of a small subset of RFCs, to understand the various sources of delay in their publication. Our work mainly builds on that by Nikkhah et al. [5], by statistically exploring RFC adoption. We build on this work with improved modelling, and by incorporating additional features that characterise not only the RFCs themselves, but the process that produced them.

## 4 CONCLUSIONS

RFC publication rates have slowed and RFCs tend to take longer to produce. Further, although the authorship pool is becoming more diverse, there are still significant issues with representation from certain regions, including Africa. From these observations, we built a predictive model to understand the features that may determine RFC success. Our results indicate that RFCs that are scalable and have clear scope tend to have better chances of deployment.

## 5 ACKNOWLEDGEMENTS

This work is supported by the UK Engineering and Physical Sciences Research Council, under grants EP/S033564/1 and EP/S036075/1.

## REFERENCES

[1] S Benthall. 2015. Testing Generative Models of Online Collaboration with BigBang (pp. 182–189). In *Proceedings of the 14th Python in Science Conference.*

[2] Christian Huitema. 2021. Evaluation of a Sample of RFCs Produced in 2018. RFC Editor. RFC 8963.

[3] Stephen McQuistin, Mladen Karan, Prashant Khare, Colin Perkins, Gareth Tyson, Matthew Purver, Patrick Healey, Waleed Iqbal, Junaid Qadir, and Ignacio Castro. 2021. Characterising the IETF Through the Lens of RFC Deployment. In *Proceedings of IMC 2021.*

[4] Heiko Niedermayer, Nikolai Schwellnus, Daniel Raumer, Edwin Cordeiro, and Georg Carle. 2017. Information Mining from Public Mailing Lists: A Case Study on IETF Mailing Lists. In *International Conference on Internet Science.* Springer, 301–309.

[5] Mehdi Nikkhah, Aman Mangal, Constantine Dovrolis, and Roch Guérin. 2017. A statistical exploration of protocol adoption. *IEEE/ACM Transactions on Networking* 25, 5 (2017), 2858–2871.

[6] Dave Thaler and Bernard Aboba. 2008. What Makes for a Successful Protocol? RFC Editor. RFC 5218.

[1] `https://www.arkko.com/tools/stats.html`