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
Information Technology University

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. This paper presents a large-scale empirical study of IETF activities, with a focus on understanding collaborative activities, and how these underpin the publication of standards documents (RFCs). Using a unique dataset of 2.4 million emails, 8,711 RFCs and 4,512 authors, we examine the shifts and trends within the standards development process, showing how protocol complexity and time to produce standards has increased. With these observations in mind, we develop statistical models to understand the factors that lead to successful uptake and deployment of protocols, deriving insights to improve the standardisation process.

CCS CONCEPTS

• **Social and professional topics → User characteristics**; • **Networks → Network protocol design**.

1 INTRODUCTION

Protocol standards are crucial to the successful operation of the Internet. A successful standard provides a basis for interoperability between systems developed by competing vendors, and supports the growth of an open ecosystem of products and services. Further, the process by which network protocol standards are developed, comprising multiple rounds of open feedback and review, has proven remarkably effective in designing high-quality and robust protocols, many of which see widespread deployment and use. Understanding the Internet standards development process, and how it produces successful protocols is, therefore, important if we are to understand the Internet and how it has evolved.

One of the main organisations that develops protocol standards is the Internet Engineering Task Force (IETF). The IETF was founded in 1986, following on from the US Government-funded effort that developed the early Internet. It has since grown to become a global community of network protocol designers, vendors, network operators, and researchers that develop and publish open network protocol standards and operational guidelines. The IETF publishes its standards, and other documents, in the RFC series (https://www.rfc-editor.org). This series comprises around 9,000 documents, authored over 50 years, and provides a rich history of the development of the Internet and its protocols [9].

While the standardisation process, taken as a whole, has clearly been successful, there are many RFCs that do not see widespread deployment. Understanding the reasons for this is complex. The success or failure of a protocol specified in a particular RFC may depend on factors beyond its technical quality. Standardisation is an inherently social and political process [5, 15], requiring cooperation and consensus among a growing number of stakeholders. For example, in 2020, IETF contributors submitted 7,547 draft documents, sent 118,537 emails to 335 mailing lists, participated in 3 plenary meetings, 256 interim meetings, and produced 309 RFCs. However, while the process has evolved and scaled, it has also slowed, with each RFC taking on average 1,170 days from its first draft to publication in 2020, an increase from 469 days in 2001.

With this growing complexity in-mind, we argue that it is vital to gain a coherent understanding of the activities that take place within the IETF, as well as the key factors that may predict the success of a protocol standard.

Using a dataset covering 8,711 RFCs, 4,512 authors and 2,439,240 emails (described in §2), we characterise the key standardisation activities within the IETF. We first focus on the RFCs being published (§3.1), to reveal that publication rates have slowed and that recent RFCs have tended to take longer to produce with more preceding draft versions. Recent RFCs also have more complex dependencies on other RFCs, and impose a greater number of normative requirements on implementers than earlier documents, suggesting wider integration across documents and stakeholders. We then consider the authors that produce these RFCs (§3.2), finding that the authorship pool is becoming more geographically diverse, with an increasing proportion of authors from Europe and Asia. We then explore the email-based discussion and interactions that underpin much of the standards development process (§3.3). We find that email volume has grown, even as RFC production has slowed.

Key findings include from analysis of this dataset include:

- Standardisation is taking longer and standards are more complex. More drafts are produced for each RFC, and RFCs update, obsolete, and cite an increasing number of drafts and RFCs. We argue that efforts need to be taken to better facilitate the collaboration process.
- The proportion of RFC authors that are from North America is declining, with an increase in authors from Europe and Asia. The IETF has signalled a desire to increase geographical diversity [1, 8]. While there is no stated goal, we found that Africa and South America continue to be underrepresented in the authorship pool, in comparison to their population sizes.
- Companies like Huawei and Google have become more influential in recent years, while contributors like Nokia and Microsoft have been in decline and have become less relevant in the standardisation process.
- Email volume has plateaued at around 130,000 messages per year, despite fewer RFCs being produced. This means that, on average, more e-mails are generated for each RFC. This underlines the growing complexity of the authorship process.

Finally, we consider what factors are the most influential in determining the success of RFCs. To do this, we further derive a set of features from our exploration of the dataset, and develop a statistical model that characterises which are most influential in predicting deployment and use of RFCs (§4). We obtain promising results, with a top F1 score of 0.822. In particular, and expanding on related work (§5), we find that RFCs that build on existing work and have well-defined requirements and applicability tend to be more successful. We conclude (§6) by noting that our findings can be used by chairs to inform their authorship and publication strategies within working groups.

To facilitate and support further research, we have made our tooling and data access scripts available (§2) to allow others to reproduce and expand on our work.

2 BACKGROUND & DATASETS

We start by presenting an overview of the IETF, and the publication process for RFCs, before outlining the data sources we use within this paper. We also highlight the ethical considerations of accessing and processing this data.

2.1 IETF Primer

IETF. The IETF is an open standards organisation, which develops Internet standards via contributions and collaborations across a number of voluntary stakeholders, including academics, consultants, industry representatives, governments, and civil society organisations. Through extensive collaboration across contributors, the IETF, and associated organisations, develops Internet standards and other documents. These are published by the RFC Editor (`https://www.rfc-editor.org`) in four *publication streams*: the IETF stream, the Internet Research Task Force (IRTF) stream, the Internet Architecture Board (IAB) stream, and the Independent Submission stream. There is also a fifth, legacy stream, comprising RFCs published prior to the adoption of separate publication streams in July 2007 [11]. While the IETF is an open standards forum that develops technical standards and operational guidelines for the Internet, the IRTF is an associated organisation that promotes longer-term research, and the IAB provides long-range technical direction for Internet development. The Independent Submission stream "allows RFC publication for some documents that are outside the official IETF/IAB/IRTF process but are relevant to the Internet community" (`https://www.rfc-editor.org/about/independent`).

RFC creation process. The standards development process is an inherently collaborative activity. Most day-to-day work is conducted on public mailing lists, in conjunction with three plenary meetings and numerous interim working group meetings per year. The mailing lists are broadly split into three categories: announcement lists, where replies are not allowed; non-working group lists, for discussing topics that do not relate to the work within an IETF working group or IRTF research group; and working group and area lists, where technical discussions take place.

The process of RFC publication begins with the submission of an *Internet-Draft*. Whereas anyone can post a draft, not all drafts become RFCs. After a draft is first posted, multiple revisions might take place resulting in multiple versions of the draft. Each new draft is announced on one or more mailing lists related to the topic of the draft, soliciting feedback and encouraging discussion. Drafts are initially posted by individuals. For publication under the IETF stream, drafts must then be adopted by a working group, where, via further revision, the draft may ultimately be published as an RFC. The process of managing drafts, and the degree and type of peer review conducted prior to their submission to the RFC Editor for publication, differs between streams.

Once the technical development of the draft is complete its publication is managed by the RFC Editor, who maintains the master archive of the RFC documents, along with an index of metadata pertaining to the RFCs and their authors. Finally, once an RFC has been published, deployment is voluntary, and therefore not all RFCs are widely implemented.

2.2 Data Sources

We rely on a number of data sources to conduct our study.

RFC Editor. The RFC Editor maintains an index of all RFC publications, alongside metadata related to each document, *e.g.*, authors, affiliations. We gather all entries for RFCs published through the end of 2020, giving a total of 8,711 RFCs.

Datatracker. Much of the work of the IETF, IRTF, and IAB is managed using the Datatracker (`https://datatracker.ietf.org`). The Datatracker is an administrative database, containing information about contributors, working groups, and meetings. In addition, the Datatracker is used to manage Internet-Draft submissions. This provides a comprehensive source of metadata about authors, and the evolution of drafts as they work their way through the standardisation process. The Datatracker was introduced in the early 2000s, and contains little historical data about RFCs produced before its creation. We have extracted relevant metadata from the Datatracker, using its REST API, for all RFCs published since 2001. This gives us data pertaining to 4,512 authors, as well as richer metadata for 5,707 RFCs. In the following sections we will describe the exact information contained within the Datatracker.

Email archives. The IETF maintains reasonably complete email archives relating to working group discussions, meeting, and other activities (`https://mailarchive.ietf.org`). Using the public IETF IMAP server, we gather all available messages contained within this archive, covering 2,439,240 messages, sent from 74,646 unique email addresses, across 1,153 lists. This snapshot was taken on April 18th 2021.

Mapping emails to contributors. A key challenge is attributing each email to an individual contributor. Although each, naturally, contains a `From` field, we find that some contributors use multiple addresses. Beyond this, it is necessary to map each email to the respective person in the Datatracker and RFC Editor datasets. Thus, we perform entity resolution on email senders, and match email senders with their Datatracker profiles. We assign each sender a unique *Person ID*, which is associated with a set of name and address variations from the Datatracker.

Entity resolution and Datatracker matching takes place in multiple stages. First, we check if the sender's email address has a Datatracker profile. If so, we associate all of their Datatracker metadata with a person ID, and label the message as having been sent by that ID. Next, if an email's sender does not appear in the Datatracker, we check, based on their name, if they have already been assigned a Datatracker record. If so, the message is labelled as having been sent by that ID, and the set of names and addresses associated with that ID is updated to include the email's sender name and address.

These two stages – matching with the Datatracker, and merging previously seen names and addresses – accounts for the majority (60%) of messages. If an email's sender is not in the Datatracker, and the name and address has not been previously seen, a new person ID is generated. This accounts for a small portion (10%) of messages. This is reasonable, given that Datatracker profiles have become necessary for many of the IETF's day-to-day activities.

As a final processing step, we label each person ID as either being a *contributor*, *role-based*, or *automated*. A contributor ID refers to any standard user participating in the IETF; role-based IDs reflect addresses that are used by the holders of a particular organisational role, such as the IETF chair; and automated IDs are those addresses that are system-specific, such as GitHub or IETF notification and announcement addresses. Role-based and automated IDs account for the remaining 30% of messages.

Note, within the mail archive, spam-indicating headers are present for most messages since 2009, and we confirm pre-filtering is performed by the IETF mail servers. As an extra validation step, we ran a spam filter (SpamAssassin, `https://spamassassin.apache.org`) over all the messages. Both sources indicate there is very little spam (less than 1%), so it should not have significant impact on our findings.

Manually labelled dataset. We also use data collated by Nikkhah et al. [13]. This dataset labels the "success", alongside other features like scope and value, of 251 RFCs. 155 of these also appear in our set of RFCs that have Datatracker metadata available. We later use this to explore the features that best predict the success of an RFC.

Microsoft Academic citation data. We use the Microsoft Academic Knowledge API to identify indexed academic work that cites each RFC. We use the Microsoft Academic Graph instead of other sources (*e.g.*, Google Scholar) as it provides time-stamped citations.
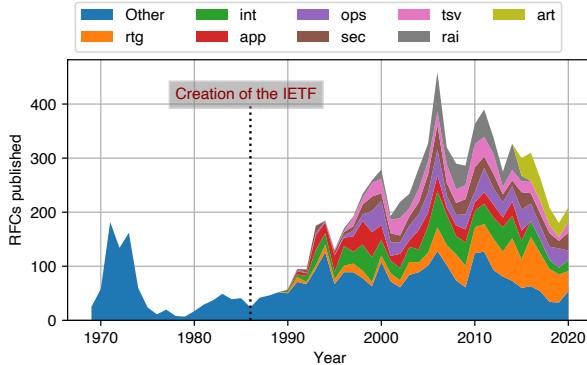
Reproducibility and data access. We have developed a library, *ietfdata*, that fetches the RFC Index, communicates with the IETF Datatracker's open API, and retrieves messages from the IETF's IMAP mail archives. This library appropriately regulates access, caches data to minimise the impact on the infrastructure, and performs necessary post-processing. This paper makes use of v0.5.0 of the ietfdata library, available from `http://dx.doi.org/10.5525/gla.researchdata.1188`. Ongoing development of this library is coordinated on GitHub at `https://github.com/glasgow-ipl/ietfdata`. All code, including for analysis and plotting, is available at `http://dx.doi.org/10.5525/gla.researchdata.1187`.

2.3 Ethical Considerations

The data we analyse is extracted from public IETF archives and APIs. We have taken steps to ensure ethical compliance. To ensure that our access to these services does not cause operational problems for the IETF, we are in regular contact with the IETF Tools Team and Secretariat, as well as the operators of the Datatracker and mailing list archive. We have extensively discussed our work with IETF leadership (IETF and IAB Chairs, the IETF Executive Director, and the IRTF Chair, who is a co-author on this paper) to ensure that our access falls within their acceptable use policies.

Participation in the IETF is dependent on agreement to abide by the policies and procedures described at `https://www.ietf.org/about/note-well`, including the privacy policy at `https://www.ietf.org/privacy-statement`. These make explicit provision that mailing list archives and the metadata contained in the Datatracker system will be made public, and it is this public data that we process to extract the aggregate statistics presented in Sections 3.1, 3.2, and 3.3, as well as the features analysed in Section 4. We transfer and store data securely, and retain it only for the time needed to perform the analysis. Since we operate entirely using the public APIs provided by IETF, we have no access to private data about individuals.

In balancing these ethical considerations with the reproducibility of our work, we provide the tools needed to access the datasets from the relevant IETF sources, rather than the data itself.

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


Figure 2: Number of publishing working groups. *Other* includes legacy RFCs, IRTF research groups, and non-IETF streams.

3 TRENDS AND CHARACTERISATION
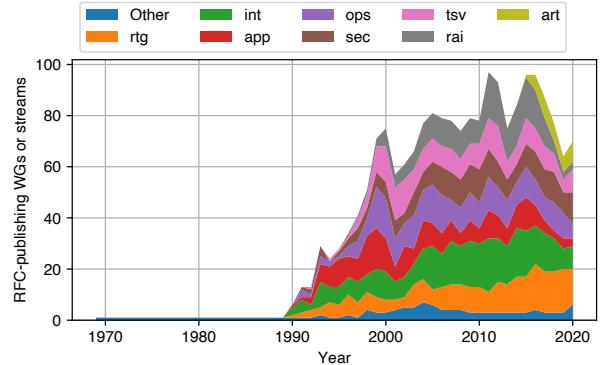
In this section, we analyse our three main datasets, looking at RFCs, their authors, and the mailing list interactions that underpin the standardisation process. The data characterisation undertaken in this section will provide the basis for the features that we model later in Section 4.

3.1 Exploring RFCs
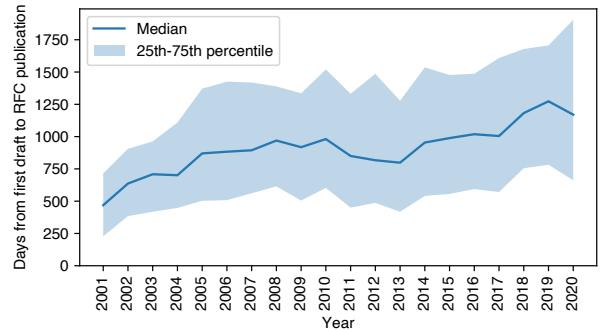
Growth of RFCs. In total, 8,711 RFCs have been published through to the end of 2020. Figure 1 shows how publication trends, in terms of IETF areas and non-IETF streams, have changed over time. We identify three broad publication phases in the RFC series. First, in 1969 through 1974, RFCs are published at a rapid rate during the initial development of the ARPANET. Then, from 1975 through 1985, development slows. This reflects a relatively small community that is gaining real-world experience with the network and developing a small core of applications and protocols. Finally, with the creation of the IETF and the introduction of the National Science Foundation Network (NSFNET) in 1986, both the community, and the number of RFCs published, starts to expand rapidly. This is further driven by the opening of the network to commercial and public use in the mid-1990s, and continues to this day.

As shown, the annual RFC publication rate was highest in 2005, at the peak of the standardisation efforts for SIP and related standards for voice-over-IP and Internet telephony. The rate of publication has slowed in recent years, following the completion of large work programmes relating to HTTP/2 and WebRTC.

Role of Working Groups (WGs). From the creation of the IETF in 1986, the growing community with its interests in an increasing set of applications and protocols, has been split into working groups. These working groups are chartered with a focus on a well-defined programme of work, and exist within areas that have a broader focus. As shown in Figure 1, the output of different areas has remained relatively stable over time. The most notable trends begin with the creation of the Real-time Applications and Infrastructure (rai) area from within the Transport (tsv) area, and its later merger with the Applications (app) area to become the Applications and Real-Time (art) area around 2014. Additionally, we also observe the


Figure 3: Days from first draft to RFC publication

significant growth in output of the Routing (rtg) area, owing to the ongoing development of standards for MPLS, service function chaining, and fat tree routing in data centres.

To give a sense for the broader productivity of the IETF, Figure 2 shows the number of working groups that publish RFCs each year. This highlights how the structure of the IETF has grown to accommodate its larger community: in the early 1990s, fewer than 20 working groups were actively publishing RFCs, while in recent years there has typically been at least 60 different publishing groups, with a peak of 97 active working groups, and other activities, in 2011.

Timeline of RFCs. While the IETF has clearly seen rapid growth since its creation, its output has recently stabilised, or even begun to decline. To explore these trends further, we look at the process and documents behind the RFCs themselves. Since 2001, much of the IETF's day-to-day activity has been managed using the *Datatracker*, which stores metadata about Internet-Drafts, meetings, and people. To further explore the RFC series, we make use of metadata gathered from the Datatracker. This metadata is only available for RFCs published since around 2001: in total, we explore the metadata for Internet-Draft leading to 5,707 RFCs.
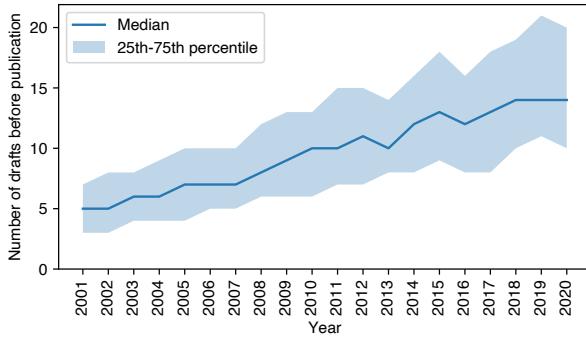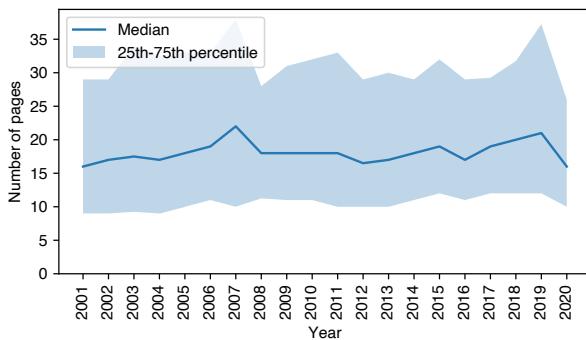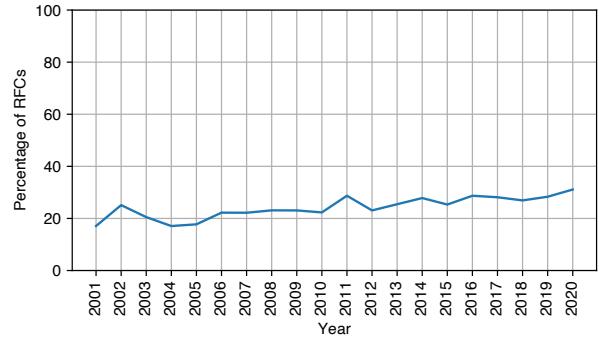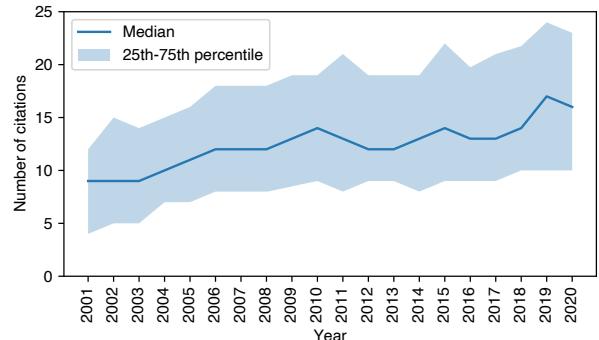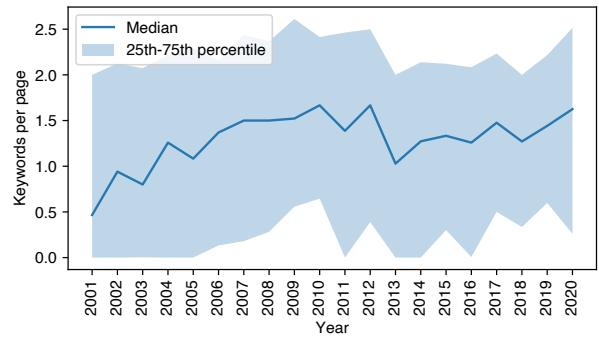
Figure 4: Number of drafts per RFC

Figure 5: RFC page counts

Figure 6: RFCs that update or obsolete previous RFCs

Figure 7: Citations from RFCs to other Internet-Drafts and RFCs per RFC

Figure 8: Keyword occurrences per page

Figure 3 plots the median number of days from the submission of an RFC's first draft, through to its publication as an RFC. This shows a clear trend: RFCs are taking longer to make their way through the standardisation and publication process. The median number of days to publication was 469 in 2001, rising to 1,170 in 2020. Further, Figure 4 shows the median number of Internet-Drafts that are posted before an RFC is published. Days to publication and number of drafts are strongly correlated, suggesting that the time is spent making changes to the document. This may go some way towards explaining the decline in output of the IETF: each RFC is taking longer to produce, with more revisions before publication.

RFC Length. One may conjecture that this slowdown is driven by longer RFCs that contain more material. To explore this, Figure 5 shows the median page count of RFCs. This shows that the increase in the duration of the standardisation process for RFCs cannot be attributed to RFCs becoming longer: median page counts have remained stable.

Relationships between RFCs. RFCs themselves may be becoming more complex as a result of them needing to describe their relationship with older standards. As the Internet has evolved and matured, applications and protocols must be maintained, and new standards need to interoperate with existing RFCs. Figure 6 highlights this, showing the proportion of RFCs that are published each year that update (*i.e.*, extend or augment) or obsolete (*i.e.*, replace) one or more previously published RFCs. Intuitively, this percentage has

slowly increased as the IETF has matured: in 2020, more than 30% of RFCs updated or made obsolete a previous RFC. Figure 7 expands on this, showing the median number of citations from each RFC to other Internet-Drafts and RFCs. This similarly shows that RFCs are increasingly referring to prior work.

Use of requirements-setting language. Figure 8 further confirms the growing complexity of RFCs, showing how the use of keywords has evolved over time. Keywords are used in RFCs to indicate the normative requirements an RFC imposes on implementations.

Figure 9: Citations to RFCs received within two years of publication from articles indexed by Microsoft Academic
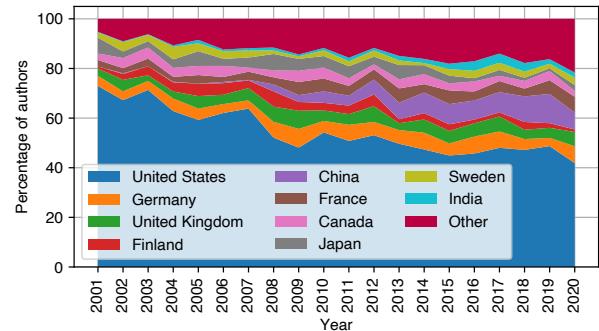
Figure 10: Citations to RFCs received within two years of publication from other RFCs
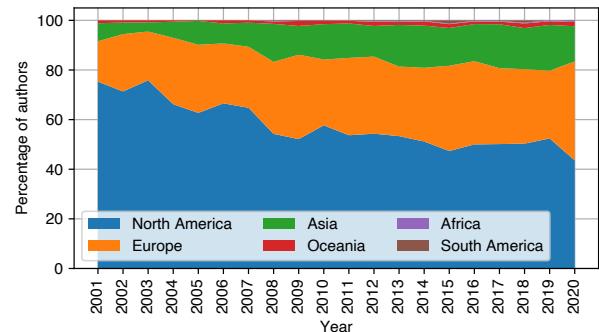
Figure 8 shows the total number of occurrences of each of the RFC 2119 [6] keywords (*i.e.*, MUST, MUST NOT, REQUIRED, SHALL, SHALL NOT, SHOULD, SHOULD NOT, RECOMMENDED, MAY, OPTIONAL), divided by the page count of the RFC. As shown, the median number of keywords per page grew from 2001 through to 2010, indicating a growing number of requirements being expressed in RFCs, before plateauing in recent years.

Academic impact of RFCs. Finally, we can measure the academic impact of RFCs, in terms of the number of citations they receive. Figure 9 shows the median number of citations that RFCs receive, by publication year, within two years after their publication, from articles indexed by Microsoft Academic. By restricting the measurement period to the two years following each RFC's publication, we are able to consistently measure citation counts. Figure 9 shows a declining trend: RFCs are seeing fewer academic citations in the two years following their publication, suggesting a fall in the academic value of RFCs. Figure 10 plots the same metric, but for citations from other RFCs. This shows a similar declining trend.

Summary. RFCs are taking longer to produce, and they go through a greater number of revisions before publication. Further, they increasingly update or reference previously published RFCs, and make greater use of requirements-setting keywords. Yet, while these measures indicate that the overall complexity of standards


Figure 11: Authorship countries (normalised)
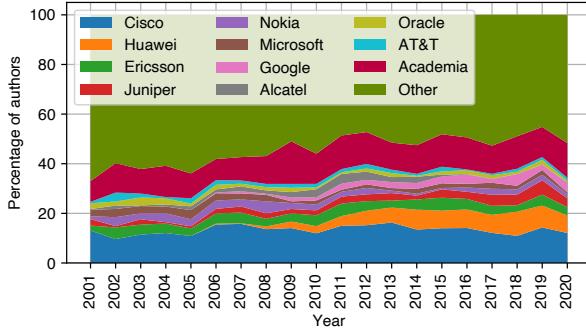
Figure 12: Authorship continents (normalised)

documents is increasing, none of these factors strongly correlate with the time taken to publish RFCs. This indicates that these factors aren't driving the increasing duration of the standardisation process. In the sections that follow, we explore trends in authorship and community interaction that may also impact the process.

3.2 Exploring Authorship

The IETF Datatracker maintains metadata about document authors, including their names, email addresses, affiliations, and location information. This dataset does not cover the entire RFC corpus, with metadata available for authors of RFCs published from 2001, and where it has been provided. Country data is available for around 70% of authors, while affiliation information is given for around 80%. In this section, we look at the authorship of all of the RFCs published each year, using the available metadata. An author is counted once in a year for each affiliation or location they hold.

Geodistribution of authors. Figures 11 and 12 show the proportion of authors from countries and continents, respectively. The IETF has signalled that it wishes to encourage greater geographical diversity [1, 8]. Without an explicit goal, we frame our findings within the context of global population distribution. We find that North America, while still disproportionately over represented, is becoming less dominant. 75% of authors were from North America in 2001, and this has declined to 44% in 2020. At the same time, representation of both Europe and Asia has grown, from 17% to 40% and 6% to 14%, respectively. However, Africa and South America

Figure 13: Authorship affiliations (normalised)

Figure 14: Academic affiliations (normalised)

Figure 15: Percentage of new authors per year.

remain heavily underrepresented, with only ≈0.5% of authors coming from either continent in 2020. This suggests that, if the IETF is to become more geographically representative, further efforts are needed.
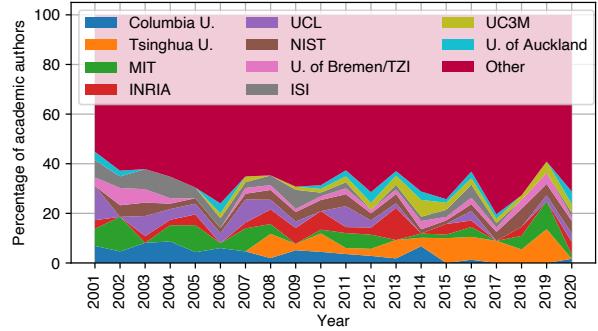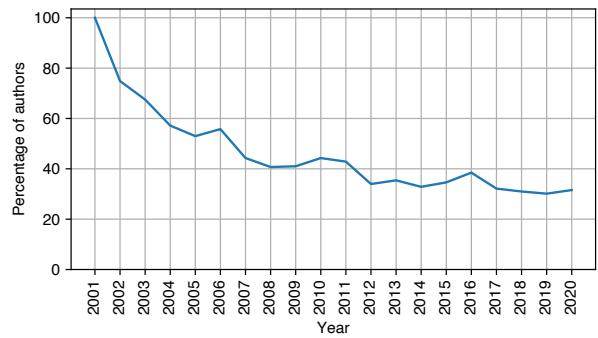
Affiliations. Figure 13 shows the top 10 affiliations by proportion of authors each year. Affiliation data is gathered from the Datatracker, and processed to normalise affiliation names, removing common variations in spelling, and amalgamating known subsidiaries and merged companies. For example, Huawei and Futurewei are combined as Huawei, and Sun Microsystems is merged with Oracle.

From Figure 13, we observe several interesting trends. First, 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 being affiliated with it in 2020.

We also observe the decline of a number of affiliations. Microsoft and Nokia, having peaked with 3.3% and 3.6% of authors, had 0.7% and 1.7% of authors in 2020, respectively, with the absolute number of authors from both companies also declining.

That we are able to observe changing trends in author affiliations, and in particular, that participants from new companies contribute, indicates that the IETF has broader relevance beyond a narrow set of companies. While we do not demonstrate any causal link, it is encouraging that commercially successful companies opt to enable their employees to actively participate in the IETF. Care is needed to ensure that this relevance is maintained, however: we observe that the author pool has grown less diverse in terms of companies that are represented. 35.4% of authors came from the top 10 affiliations in the dataset in 2020, compared with 25.6% in 2001.

Academia and consultants. Academic affiliations are those where the affiliation name contains "University", "Institute", or "College", while consultancy affiliations are those that contain "Consultant". Affiliation data has been normalised to remove common abbreviations (*e.g.*, "U." for "University") and to translate non-English affiliations. As shown in Figure 13, we find that an increasing number of authors come from academic affiliations, growing from 8.1% of authors in 2001, to 13.6% in 2020, having peaked at 16.5% in 2009.
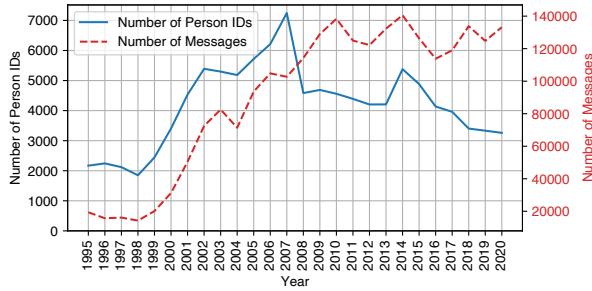
The number of consultants, by our measure, has remained stable, accounting for 2% of authors in 2020. The remaining authors are largely from industrial affiliations.
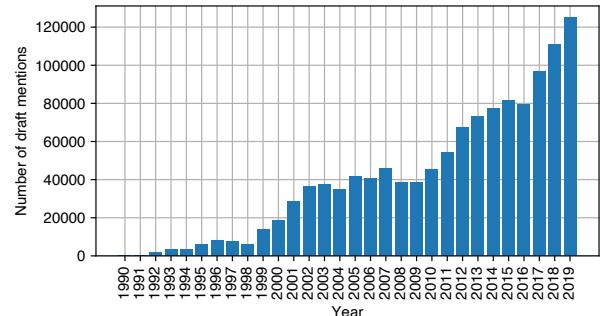
Figure 14 shows the top 10 academic affiliations in the dataset, and the percentage of academic authors that have those affiliations over time. In general, academic affiliations are each typically held by a small number of authors. We can see a number of trends in academic authorship, with fewer authors from Columbia University, MIT, and ISI in recent years, and the rise of Tsinghua University and University Carlos III of Madrid.

Arrival of new authors. Finally, Figure 15 shows the percentage of authors in each year that have not previously authored an RFC. Given that the dataset used here begins in 2001, 100% of authors are new in that year. The more stable trend in recent years likely highlights the churn in RFC authorship, with around 30% of authors each year having never previously authored an RFC.
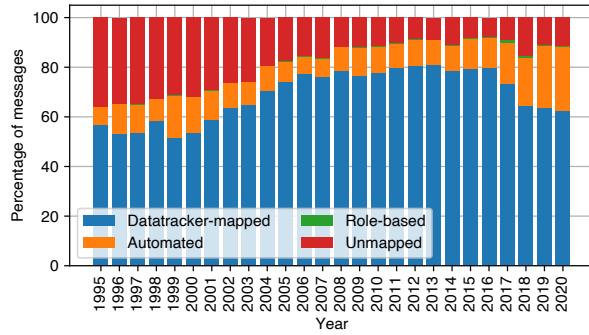
Summary. The trends highlighted in this section indicate a pool of authors that is diversifying and changing over time, with a growing proportion of authors from outside of North America, contributions from new companies and affiliations, and relatively high authorship churn. Despite this, we find that certain regions and groups are not well represented in the IETF (*e.g.*, contributors from Africa), and that the authorship pool is becoming increasingly centralised, with a third of authors coming from the top 10 affiliations. This suggests,

Figure 16: Number of Person IDs exchanging emails each year


Figure 17: Number of messages exchanged per year across different categories: messages mapped to the Datatracker (Datatracker Person-ID); messages by automated email addresses (Automated); messages by role-based addresses (Role-based); and messages not mapped to the Datatracker (New Person-ID)

if the organisation is concerned with being more representative, further efforts are needed.

3.3 Exploring Email Interactions

The publication of RFCs is largely underpinned by mailing list interactions, which are used to discuss and finalise RFCs. Our data, which starts from 1995, confirms their vital role, with 1,153 mailing lists, containing 2,439,240 emails from 74,646 unique email addresses.

Volume of emails. Figure 16 presents the number of emails sent across the last 25 years, showing that email volumes have grown significantly with time, plateauing at around 130,000 annual messages since around 2010. The figure also shows how the number of unique Person ID counts (as described in §2) observed every year in the mail lists, showing a decreasing trend in the number of contributors.
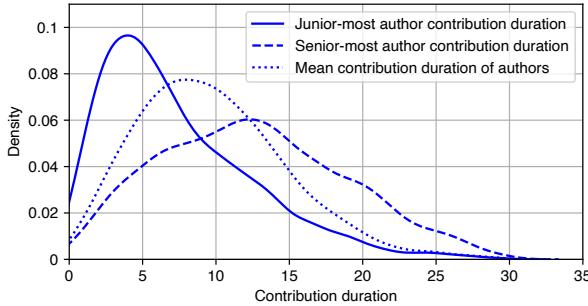
Figure 17 shows the relative email volume by type. We see that an increasing fraction of messages originate from automated addresses. This is largely as a result of the growing use of GitHub, and other version control systems, for managing drafts. There were 122 active IETF working groups at the time of writing: of these, 17 listed a GitHub repository in their metadata. The QUIC working group, as

Figure 18: Number of draft mentions in each year found in the mail lists

one example, has replaced the typical email list discussions with GitHub issues: indeed, this is a significant part of the surge observed in 2016. Given that most working groups use mailing lists to manage their activity, we do not further analyse interactions that take place on GitHub in this paper. As a result, Figure 17 understates the volume of interactions: the plateau observed in recent years is at least somewhat attributable to the shift to GitHub and similar services. As this shift continues, it will become important for future work to consider these interactions.

Discussion of drafts. The mailing lists are, other than plenary or interim meetings, the primary means for the discussion of drafts. To measure how often drafts are discussed, we next identify mentions of drafts in mailing list messages. We therefore extract any mention of a draft (beginning `draft-`) or RFC (*i.e.*, , "RFC" followed by a number). Figure 18 presents the number of drafts mentions in the emails per year. Separate mentions of the same draft are counted as different mentions, as we want to observe the entire volume of mentions. We observe a strong increase in the number of mentions over time. This is largely driven by the growing number of drafts being published. In fact, we find a Pearson correlation of 0.89 between the number of drafts published and the number of mentions. This speaks to the influence of emails that mention drafts in driving draft production.

Contribution duration. We now look into the longevity of contributors, in terms of the number of years that they actively participate in any IETF mailing lists. We define the *contribution duration* of a participant as the length of time that they have contributed to the mailing lists. To do this, we look at the contributors who first contribute to mailing lists between the years 2000 to 2013. We limit our analysis to 2013, since the longevity of contributors that first participate more recently cannot be determined. For each year, in the period 2000 to 2013, we look at those who first contribute in each year and the number of years that they then go on to remain active in the mailing lists. For example, a participant who first sent an e-mail to an IETF list in 2010, and last sent an e-mail to an IETF list in 2018, will have a contribution duration of 9 years.

We generate Gaussian Mixture Models for observing different clusters of the maximum duration that contributors remain active for. These models identified three broad clusters that contributors can be categorised within:

Figure 19: Contribution duration distribution of authors of RFCs: junior-most author of each RFC, senior-most author of each RFC, and mean contribution-duration of all authors of each RFC

- *Young contributors* who leave within 1 year of joining.
- *Mid-age contributors* who go onto remain associated for more than 1 year, but less than 5 years.
- *Senior contributors* who go onto remain associated for 5 or more years.

Interactions based on contribution duration. We use the three contribution duration categories defined above to characterise the interactions between contributors belonging to different categories, as well as the authors of each RFC. We look at outgoing and incoming interactions with RFC authors in the period between the first draft and publication of the RFC. If this period is less than two years, we look at the activities of authors for two years before the RFC was published. Interactions are defined from the viewpoint of the RFC authors:
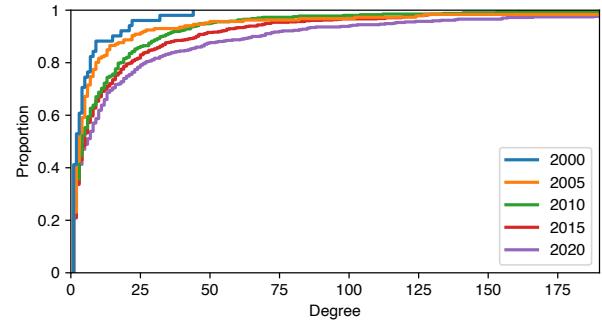
- *Outgoing interaction*, where an RFC author responds to an email from other contributors (*i.e.*, email sent).
- *Incoming interaction*, where contributors respond to an email by the author (*i.e.*, email received).

For each interaction type, we count the number of messages and number of unique contributors involved in those interactions. For each RFC, we analyse the interaction activities based on three measures of author contribution duration:
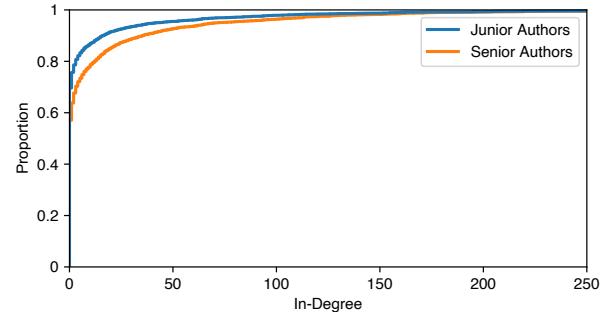
- *Junior-most author*, *i.e.*, the author with the lowest contribution duration at the time of publication.
- *Senior-most author*, *i.e.*, the author with the highest contribution duration at time of publication.
- *Mean contribution duration* of all of the authors of the RFC at the time of publication.

Figure 19 shows the distribution of contribution duration of each of these three measures. This shows that the majority of junior-most authors have participated for less than 5 years in the IETF, whereas the majority of senior-most authors have participated in the IETF for more than 10 years. In fact, 35% of authors exceed 15 years of IETF participation, showing that RFCs tend to be authored by a mix of seniority levels.

Evolution of interactions. Figure 20 shows the gradual drift in annual degree (*i.e.*, the number of people interacted with) of RFC authors. The degree of authors has substantially increased over

Figure 20: CDF showing drift in annual degree (interaction with their network) of RFC authors over the period 2000-2020.

Figure 21: CDF showing number of senior contributors sending messages (in-degree) to junior and senior authors.

time. For instance, in the year 2000 only 5.5% of the authors had a degree of over 25, whereas by the year 2015 almost a quarter of the authors had a degree over 25. This confirms that, on average, more recent RFCs generate greater discussion. This may explain increasing publication times: authors spend more time interacting with other participants, and these discussions are likely to lead to more drafts.

We can also contrast the interaction patterns of the junior vs. senior authors. Figure 21 presents the CDFs of the in-degree across both junior and senior authors. It shows that the incoming interactions from senior contributors to junior authors are significantly less than the incoming interactions from senior contributors to senior authors. Nearly 55% of junior authors receive messages from fewer than 10 senior contributors, whereas nearly 65% of senior authors receive messages from more than 10 senior contributors. This speaks to the differing roles played by these sub-populations: senior authors act as hubs through which substantial volumes of interaction flow.

Summary. Emails play a vital role in underpinning RFC publication with approximately 130,000 emails sent per year. We have shown that the seniority of both participants and RFC authors fundamentally changes the volume of interactions that they have. These trends are likely to have implications for the IETF community, especially as it tries to encourage new participants.

4 EXPLORING RFC SUCCESS FACTORS

The trends that we observed in Section 3 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, with differences in how senior and junior authors and contributors interact. 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.

4.1 Overview of Approach

We build on the work of Nikkhah et al. [13]. However, while the focus of that paper was on a set of features that were derived from the standards documents themselves, we augment this with a set of features derived from the author and email interaction datasets described in the previous section.

To explore the factors that influence protocol deployment, we follow a multi-step process:

Step 1. We reproduce the logistic regression model in [13], to obtain similar results (as shown in Table 3). This paper published an expert annotated dataset, labelling RFCs as "successfully deployed" or not. This pertains to whether the RFC was implemented in-the-wild and saw widespread (largely commercial) uptake. This dataset covers 251 RFCs published between 1983 and 2011, and includes 20 features. We create a baseline logistic regression model with $F_1 = 0.762$, $AUC = 0.650$. The results are similar to those reported in [13] ($AUC = 0.670$).

Step 2. We curate a set of additional features (§4.2), driven by our findings in Section 3. These are derived from the RFCs, their authors, and the mailing list interactions. Given that these features are only available for RFCs where Datatracker metadata is available, we model these features across a subset of the original labelled RFCs from [13], looking only at the 155 RFCs where all features can be calculated. We will show that modelling these RFCs (§4.3) with the expanded feature set substantially expands the predictive power of the (Step 1) baseline logistic regression model, with $F_1 = 0.820$, $AUC = 0.822$.

Step 3. We then train classification models to predict if an RFC will be deployed, using our expanded feature set. We test the model with the manually labelled RFCs and show that we can successfully classify RFCs as deployed or not, with a decision tree-based model having $F_1 = 0.822$, $AUC = 0.838$.

4.2 Classification Features

We begin with the set of features originally derived by Nikkhah et al. [13]:

(1) *Area*: Applications and Real-Time (ART), Internet (INT), Operations and Management (OPS), Routing (RTG), Security (SEC), or Transport (TSV);
(2) *Scope*: Local, End-to-End (E2E), Bounded (BN), or Unbounded (UB);
(3) *Type*: New (N), new with incumbent (NI), Backward compatible extension (EB), or Extension (E);
(4) Change to others (CO);

(5) Scalability (SCAL);
(6) Security (SCRT);
(7) Performance (PERF);
(8) Adds value (AV); and
(9) Network effect (NE).

We refer the reader to [13] for a full explanation of each of these feature. We further define an additional set of document-based features, derived from our characterisations in §3.1:

(1) Days from first draft to RFC publication;
(2) Number of drafts before RFC publication;
(3) Number of outbound citations to RFCs/Internet-Drafts;
(4) Page count;
(5) Number of inbound citations from Microsoft Academic articles, one and two years after publication;
(6) Number of inbound citations from other RFCs published within one and two years after publication;
(7) RFC updates or obsoletes a previous RFC (yes/no);
(8) Keywords per page; and
(9) Topics: we use Latent Dirichlet Allocation (LDA) [4] to induce 50 topics on the texts of all existing RFCs, and use the 50-dimensional probability distribution over topics for a given RFC as the feature vector.

Next, we derive a set of author-based features, from the work described in §3.2:

(1) Number of authors on the RFC;
(2) If at least one author of previously published RFC (yes/no);
(3) If at least one author in North America, Europe, or Asia (yes/no);
(4) If at least one author from Cisco, Huawei, or Ericsson (yes/no);
(5) Authors with diverse affiliations (yes/no);
(6) Authors in more than one continent (yes/no);
(7) At least one academic author (yes/no); and
(8) At least one consultant author (yes/no).

Finally, we derive a set of features based on email interactions, from §3.3:

(1) Number of emails mentioning the Internet-Drafts that precede the RFC's publication;
(2) Mean number of emails sent to all RFC authors, for each sender contribution-duration category (young, mid-age, senior);
(3) Mean number of contributors sending emails to any RFC author, for each contribution-duration category;
(4) Number of emails sent to most junior and senior RFC authors, for each sender contribution-duration category; and
(5) Number of contributors sending emails to most junior and senior RFC authors, for each contribution-duration category.

4.3 Modeling Methodology

With our expanded feature set, we next build a classifier to predict the success (*i.e.*, deployment) of RFCs.

Feature engineering. Our expanded feature set, including all variants of each feature, is very large, totalling 177 features. Given that our additional features can only be calculated for a limited number of data points (155 RFCs), it would be infeasible to conduct statistical analysis, and to build a predictive model, as there are

Feature Name	Coef.	P>|Z|
Change to others (CO)	0.0001	1.000
Adds value (AV)	**0.7828**	**0.009**
Security (SCRT)	0.3830	0.253
Scalability (SCAL)	**0.8755**	**0.100**
Performance (PERF)	0.5108	0.323
Microsoft Academic citations, 1 year	0.2380	0.234
Updates others (Yes)	0.2877	0.514
Obsoletes others (Yes)	**1.5315**	**0.001**
Keywords per page	**0.3409**	**0.083**
Inbound RFC citations, 1 year	0.6112	0.011
Junior-author → Senior (messages)	0.1226	0.463
Young → Senior-author (messages)	0.2168	0.244
Senior → Senior-author (people)	**-0.2902**	**0.096**
-00 draft mentions	-0.2198	0.187
Final draft mentions	0.1469	0.370
All draft mentions (normalised)	-0.0504	0.755
-00 draft mentions (normalised)	-0.1052	0.525
Author count	-0.0941	0.561
Days to publication	0.1560	0.340
Draft Count (DC)	0.1845	0.262
Outbound citation count	0.2256	0.173
Page count	**0.3468**	**0.054**
Topic 13 (MPLS)	**-0.5629**	**0.068**
Topic 19	-1.2596	0.111
Topic 31	**-2.0698**	**0.021**
Topic 44	0.3992	0.129
Topic 45	**0.3289**	**0.097**
Area (INT)	-0.1671	0.683
Area (OPS)	0.5108	0.323
Area (SEC)	0.2231	0.638
Area (TSV)	0.5108	0.323
Type, Backward Compatible (EB)	0.3502	0.135
No incumbent	**0.6061**	**0.039**
Has incumbent	-0.2007	0.655
Scope, End-to-end (E2E)	**0.5878**	**0.035**
Scope, Local (L)	1.3863	0.215
Scope, Unbounded (UB)	**-1.0986**	**0.033**
Has author in N. America (Unknown)	-0.6061	0.232
Has author in Europe (Unknown)	0.1671	0.683
Has author in Asia (Yes)	**-0.8755**	**0.100**
Has author from Cisco (Unknown)	-0.4055	0.442
Has author from Cisco (Yes)	0.4463	0.163
Has author from Huawei (Yes)	-1.3863	0.215
Has author from Ericsson (Unknown)	0.0001	1.000
Has author from Ericsson (Yes)	0.5596	0.372
Has continent diversity (Yes)	-0.1911	0.538
Has an academic author (Yes)	-0.0870	0.835
Has a consultant author (Yes)	-0.6931	0.423

Table 1: Logistic regression w/o feature selection. Statistically significant rows ($p \leq 0.1$) are highlighted.

Feature Name	Coef.	P>|Z|
Inbound RFC citations, one year	**0.611**	**0.011**
Obsoletes others (Yes)	**0.531**	**0.001**
-00 draft mentions	-0.219	0.187
Scope, Unbounded (UB)	**-1.0986**	**0.033**
Has author in Asia (Yes)	**-0.8755**	**0.100**
Has author in N. America (Unknown)	-0.6061	0.232
Keywords per page	**0.3409**	**0.083**
Topic 45	**0.3289**	**0.097**
Has an academic author (Yes)	-0.087	0.835
Type, Backward Compatible (EB)	0.3502	0.135
Topic 13 (MPLS)	**-0.5629**	**0.068**
Topic 31	**-2.0698**	**0.021**
Topic 44	0.3992	0.129
Has continent diversity (Yes)	-0.1911	0.538
Young → Senior-author (messages)	0.2168	0.244
Topic 19	-1.2596	0.111
Draft Count (DC)	0.1845	0.262
Scope, Local (L)	1.3863	0.215
Has author from Ericsson (Yes)	0.5596	0.372

Table 2: Statistical analysis using logistic regression w/ feature selection. Statistically significant rows ($p \leq 0.1$) are highlighted.

Model	F_1	AUC	$F1_{macro}$
Most frequent class	.757	.500	.379
Baseline	.758	.616	.597
Baseline + FS	.762	.650	.610
Most frequent class	.724	.500	.379
Baseline	.670	.559	.547
Baseline + FS	.690	.620	.563
Logistic regression all feats	.728	.724	.666
Logistic regression all feats + FS	.820	.822	.789
Decision tree all feats + FS	.822	.838	.788

Table 3: Classifier scores on the entire dataset (251 RFCs, above) and those with our features available (155 RFCs), with or without feature selection (FS).

too few data points. To address this, we take a number of steps to reduce the feature space, while maintaining interpretability.

First, since the largest feature groups are the topics (50) and interaction features (54) we reduce both by applying the χ^2 test to leave only the top 5 features in each group. *Second*, we remove collinearity by using the Variance Inflation Criterion (VIF), removing all features with a VIF value above 5. *Third*, we apply forward Feature Selection (FS) to identify features of high predictive value, following [13]. Starting from an empty feature set, in each iteration of the forward procedure, we expand the feature set with the feature that provides the largest increase in the AUC score. The procedure

ends when there are no more unused features that yield score improvements over the current feature set. The final set of features with and without FS is given in Table 2 and Table 1, respectively.

Model training. Using the above feature set, we train two classification models relying on logistic regression and a decision tree in the `scikit-learn` library [14]. For assessing predictive performance of the models we use leave-one-out cross-validation, while for the final statistical analysis we fit a logistic regression model on the entire dataset and report the statistically significant coefficients (at significance level $p \leq 0.1$).

4.4 Evaluative Results

Using the trained models, we next evaluate their accuracy and explore which features are most predictive of RFC deployment.

Evaluation metrics. As evaluative metrics, we use the F_1 score and the area under the ROC curve (AUC), as in [13]. We find the standard F_1 score gives overly optimistic performance estimates due the data being skewed towards the positive class. Consequently, we also report an $F1_{macro}$ score that takes into account both the positive and negative class, reflecting performance more realistically. As a baseline comparison, we present results of our re-implementation of [13] (using their feature set).

Results. The prediction results of the models and feature sets are summarized in Table 3. Our baseline model performs at a similar level to that from [13]. We further obtain considerable performance improvements with our additional features. Furthermore, we find that additional performance can be obtained by using feature selection. The best performing model, with an F1 score of 0.822, is the decision tree trained on the entire feature set. Note, we also tested several non-linear models (neural networks, support vector machines with non-linear kernels). These attained similar or worse results as our decision tree model, and therefore we omit them due to space constraints.

Feature importance. We next investigate the most important features in predicting success. We posit that these can offer useful insight for working group chairs and RFC editors. The feature importance results are presented in Tables 1 and 2. Some meaningful patterns emerge here. We observe features such as adding value, being scalable, obsoleting other RFCs, page count, increased keyword usage, no incumbent RFC to compete against and limited scope all being positively correlated with the likelihood of an RFC becoming deployed. In contrast, we find that having a broad (unbounded) scope negatively impacts deployment. We will later discuss the importance of these observations (§4.5).

We also observe some curious trends that speak to the limitations of our dataset. The model finds that having an author in Asia is negatively correlated with deployment, though with only borderline statistical significance. Further analysis shows that only 10% of labelled RFCs (17 in total) have an author in Asia. As shown in §3.2, we see that demographics of the IETF are changing, with a recent notable increase in author representation from Asia. This might suggest that new authors need time to learn what makes a deployable protocol; equally our deployment data may be biased towards the Internet in North America and Europe. This finding requires much more exploration.

We further see that some of the topics extracted by LDA are also useful. For example, Topic 13 is characterised by a cluster of terms associated with MPLS, a widely deployed routing protocol. We find a negative correlation with this topic, but this is likely because there is a significant number of RFCs that propose modifications or additional features for the protocol, some of which do not see deployment. Overall, most of the results are in line with expectations, but more annotated data is required for further insights.

4.5 Discussion

Finally, we briefly discuss the salient outcomes from our modelling.

Building on existing work. Features associated with building on existing RFCs positively correlate with deployment. RFCs that obsolete earlier versions of the same protocol are likely to be deployed, indicating that the IETF community is good at identifying and maintaining protocols that are seeing use. This is also highlighted by the significance of the inbound RFC citation and "adds value" features. These show that RFCs that are later cited by other documents, or that add value to other protocols in the stack, are more likely to be deployed. Protocols that are extensible, and can be extended and adapted for new uses, are more likely to see deployment. The discussion in [7] on careful choice of extension points resonates; as do the implications for starting new work.

Limited scope. We find that having end-to-end scope (*i.e.*, where only the endpoints of a connection needs to implement the RFC) is positively correlated with deployment, whereas having an unbounded scope (*i.e.*, where the entire Internet may need to be updated) is negatively correlated. This indicates that well-scoped RFCs, that are cognisant of their deployment challenges, are more likely to see deployment. A recent IAB workshop [2] noted that deployment often occurs in unforeseen ways: limiting the changes needed to deploy a protocol facilities this.

Normative requirements. The number of keywords (*e.g.*, SHOULD, MUST) used per page, which shows the number of normative requirements an RFC imposes on implementations, also correlates with deployment. This suggests that well-specified requirements lead to robust, interoperable, implementations that see wide use. Unsurprisingly, features such as citation rates are also predictive of deployment.

Diversity. Notwithstanding the limitations of the dataset, we find that the majority of author demographic features are not significant. While §3.2 identified recent shifts in the demographics of authors, these do not appear to have a major influence in the deployment of the RFCs that they write.

Meeting a need. Our results broadly support the conclusions of [16], 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.

Summary. Our results show that RFCs that build on existing work, have well-defined requirements, limited scope, and meet a need have improved chances of being widely deployed. These document-based features have implications for how the IETF should design and specify protocols.

We found that the majority of our author-based features, including geographic and affiliation diversity, did not significantly

impact the deployment prospects of RFCs. Further work is needed to expand the dataset to ensure that this result holds across a larger set of RFCs.

Finally, the focus of our modelling in this paper has been on the impact of document, author, and interaction-based features on protocol deployment. However, this is only one part of the lifecycle of an RFC. It remains to consider the impact of these, and other, features on the key stages of an Internet-Draft's development towards becoming an RFC, such as working group adoption.

5 RELATED WORK

To the best of our knowledge, we are the first to perform a comprehensive statistical analysis of IETF activities. That said, there have been several prior efforts that have focused on specific aspects of the IETF. Jari Arkko, a former chair of the IETF, maintains a website (https://www.arkko.com/tools/stats.html) and tooling that provides various statistics about the IETF, including about its documents, authors and their affiliations. BigBang [3] is a Python toolkit for analysing online collaborative communities through mailing list data. Niedermayer et al. [12] discuss the challenges of working with large mailing list datasets, including conducting entity resolution where contributors have changeable pseudonyms. Our work addresses a number of these challenges by augmenting the mailing list data with metadata from the IETF Datatracker.

Huitema [10] carried out an evaluation of a small subset of RFCs, to understand the various sources of delay in their publication. The author observed that the main source of delay was the working group process, highlighting the need to better understand the dynamics of that process. In addition, the author found that citation counts did not correlate well with the adoption or deployment of specific RFCs.

Our work mainly builds on that by Nikkhah et al. [13], 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 standardisation process that produced them.

6 CONCLUSIONS

This paper explored the key standardisation activities within the IETF, through the lens of RFCs (§3.1), authors (§3.2) and email collaborations (§3.3). We have revealed a vibrant ecosystem, albeit one with a number of key challenges. For instance, 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 and South America. If the IETF is to become more representative, then further effort is required to build engagement in these regions. From these observations, we built a predictive model to understand the features that may determine RFC success (§4). Our results indicate that RFCs that are scalable and have clear scope tend to have better chances of deployment. We also found that the IETF itself does a good job of identifying successful protocols, and building upon them.

Despite this, there are still many aspects of future study. We intend to explore new modalities of interaction within the IETF, including the growing use of GitHub, as well as including data from

meetings such as minutes and agendas. We will also develop targeted interventions that can be better facilitate things like diversity within the organisation. As part of this and to foster further research, we have made all of our tooling available to the community.

7 ACKNOWLEDGEMENTS

REFERENCES

[1] Jari Arkko. 2013. IETF Blog: Diversity. https://www.ietf.org/blog/diversity. Accessed: 2021/09/27.
[2] Jari Arkko and Ted Hardie. 2021. Report from the IAB Workshop on Design Expectations vs. Deployment Reality in Protocol Development. RFC Editor. RFC 8980.
[3] S Benthall. 2015. Testing Generative Models of Online Collaboration with BigBang (pp. 182–189). In *Proceedings of the 14th Python in Science Conference.*
[4] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3 (2003), 993–1022.
[5] Scott Bradner. 1996. The Internet Standards Process – Revision 3. RFC Editor. RFC 2026.
[6] Scott Bradner. 1997. Key words for use in RFCs to Indicate Requirement Levels. RFC Editor. RFC 2119.
[7] D. D. Clark, J. Wroclawski, K. R. Sollins, and R. Braden. 2002. Tussle in Cyberspace: Defining Tomorrow's Internet. In *Proceedings of the SIGCOMM Conference.* ACM, Pittsburgh, PA, USA. https://doi.org/10.1145/633025.633059
[8] Dave Crocker and Narelle Clark. 2015. An IETF with Much Diversity and Professional Conduct. RFC Editor. RFC 7704.
[9] Heather Flanagan. 2019. Fifty Years of RFCs. RFC Editor. RFC 8700.
[10] Christian Huitema. 2021. Evaluation of a Sample of RFCs Produced in 2018. RFC Editor. RFC 8963.
[11] IAB and Leslie Daigle. 2007. The RFC Series and RFC Editor. RFC Editor. RFC 4844.
[12] 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.
[13] 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.
[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
[15] Pete Resnick. 2014. On Consensus and Humming in the IETF. RFC Editor. RFC 7282.
[16] Dave Thaler and Bernard Aboba. 2008. What Makes for a Successful Protocol? RFC Editor. RFC 5218.