

An Analysis of IETF Activities Using Mailing Lists and Social Media

Heiko Niedermayer, Daniel Raumer, Nikolai Schwellnus,
Edwin Cordeiro, and Georg Carle

Chair of Network Architectures and Services,
Department of Informatics, Technical University of Munich
{niedermh,raumer,schwelln,cordeiro,carle}@in.tum.de

Abstract. The Internet Engineering Task Force is an open organization that produces Internet Standards. In this paper we look at Twitter and IETF mailing lists to answer questions on IETF participation and social media usage and IETF reaction to societal events: Are Internet Standards discussed on Twitter? Who is involved in the process? Do external events like Snowden revelations in 2013 correlate with related IETF activities? To answer this, we look in particular at security-related activities at the IETF like in the TLS working group. With respect to the Snowden leaks, we quantify the impact in terms of increase in activity and contributors in related areas.

1 Introduction

From its beginnings as dedicated research network, only accessible by an exclusive subset of people, the Internet has developed into an important catalyst to societal development. Being open to everyone and everyone's technology mankind is still witnessing new applications and fields to apply them each day. Within these efforts different stakeholders put efforts into refining the Internet and its underlying technologies by defining new standards. Although the definitions of what can be named a standard differs vastly, all have in common to define technologies and techniques to be used in order to achieve a level of agreement that improves interoperability for some purpose.

Standards may be called de facto standard with a positive attitude or quasi standard with a rather negative attitude. Powerful stakeholders, often market leaders or exclusive groups controlling a market, may either explicitly or implicitly define standards to protect their position. An explicit standard definition is usually made by open or closed documents and implicit standards are made by designing technologies or programs that are used by a majority. By ironing or adopting those standards, the masses constitute the second step for defining a standard: adoption.

Within this setting, the Internet Engineering Task Force (IETF) and its affiliated organization the Internet Research Task Force (IRTF) are an institution that is focused on improving the Internet and related technologies [2]. The IETF

does that by means of RFCs that define standards and best practices. Formed in 1986 the IETF started as quarterly meetings of researchers funded by the US government. The meetings were opened to the general public in October 1986 and remained this way ever since [7,5]. The IETF remained supported by government funds until 1997 [5], since July 1995 the IETF is supported by the Internet Society (ISOC) [6].

Without formal membership or membership requirements, their contributors work on a voluntary basis. Anyone interested in contributing to the development of Internet protocols may participate. Despite that open policy, missing technical expertise or financial background, companies' policies, and formation of exclusive subgroups within the IETF may limit the openness. These factors may crush the will and possibility to contribute and constitute a problem for the IETF [8].

Our contribution is as follows: In this paper we want to put a flash light onto certain activity aspects of the IETF. We use social media data from Twitter as well as mailing list data from the IETF mailing lists. These mailing lists are the work horse of the IETF. With that we analyse interest into certain topics over time as well social media usage and presence of IETF standards. In Section 2, we discuss basic processes of the IETF standardisation. Section 3 contains the analysis and we conclude with Section 4.

2 Background

To coordinate the work of the people active in the IETF, the activities are organized in working groups, each one collaborating on a specific topic. We also have areas, like Internet or routing areas, that are formed by groups with similar topics. Working groups are created with a specific goal detailed in its charter, each has their own discussion mailing list and one or more working group chairs. After having fulfilled its purpose, a working group closes or is rechartered to add new goals. [4]

A reasonably complex structure is necessary to guarantee the quality of the standards created by this open organization. The first part is the Internet Society, that provides financial and legal support for the IETF and the Internet Research Task Force (IRTF). The Internet Engineering Steering Group (IESG) is composed of all area directors and the IETF chair, all chosen for a 2-year renewable term by the Nominating Committee, in a complex but publicly verifiable process [12]. The IESG is responsible to manage the process of creating Internet Standards and other IETF activities. It works by guiding the process and not by making decisions about the standards.

The Internet Architecture Board (IAB) is an oversight committee for the "architecture of the Internet and its protocols" [11]. The IAB reviews the charters of new working groups before they are created and is expected to keep an eye on the "big picture". [10]

Over their lifetime working groups will produce a number of documents, with the goal to publish them as *Internet Standards* or *Informational* documents. The Standards Process of the IETF works as follows:

A working group – after reaching rough consensus with its participants – forwards a draft document to the IESG. The IESG issues a last call for comments for all working groups in all areas and if the ideas or methods of the draft document reach rough consensus in the IETF as a whole and the IESG does not have concerns, the draft will be forwarded to the RFC-Editor for publishing. If the IESG or the IETF community does have concerns or the ideas and methods of the draft document did not reach rough consensus in the IETF, the working group will update the draft document to begin the process again, or the document is abandoned by the working group.

3 Analysis of IETF activities

The IETF follows a well defined process based open discussions. With its open architecture the IETF constitutes an eligible example to study actors of standardisation. Discussions in the IETF mostly take place on mailing lists, but also at the IETF meetings (three times per year) and interim working group meetings (scheduled when a working group needs it). Therefore, different events can be counted to determine the activity of a working group whereof mail exchange is the most continuous event. Jari Arkko [3] maintains a website with numerous statistics based on the official IETF documents. As this database only contains the official sources and the website only presents selected statistics we have built up our own database that we want to extend by further sources of social activity.

3.1 IETF mailing lists

Figure 1 shows the top 5 mailing lists IETF working groups of 2015 and their activity according to the absolute number of sent mails. Here and in following graphs we extrapolated the activity for 2016 on behalf of the first 3 month activities.

Each mailing list serves for discussion of its working group. Each group has an agenda published on the web. Shortly summarized the working groups listed in our graphs focus on the Transport Layer Security standard (TLS), the operation of Domain Name Systems (DNSop), data modeling for information exchange with network devices (netmod), deployment and operation of IPv6 networks (v6ops), maintenance and development of the Hypertext Transfer Protocol (httpbis), the Kerberos authentication protocol (krb-wg), the HTTP extension WebDAV (webdav), and mail security (ietf-dkim).

A few words on methodology: As spam mails do not reflect any activity we have filtered these in our analysis. We based the decision which mails to filter on the X-SpamScore field. This field is generated by Apache SpamAssassin [13] and present in the public mail archives. In the following top 5 selections we excluded special purpose IETF mailing lists like *iesg-agenda-dist* and *ipr-announce* as these do not reveal any insights into the interactions of individuals.

The SpamScore is generated by Apache SpamAssassin using rules that if hit, add or subtract a certain value to the score. Rule examples include message body

with mostly blank lines, containing keywords like 'valium', 'viagra' or 'million dollars' and user is in the whitelist (Full list: [14]). On the other hand, mails with subject containing 'draft-' are very unlikely to be spam as it is probably a reference to a certain draft document, same holds true for squarebrackets referring to a certain mailing list, as well as network related keywords. Additionally, we did not interpret mail that are an answer to a non spam mail (denoted by the 'RE: *' subject) as spam.

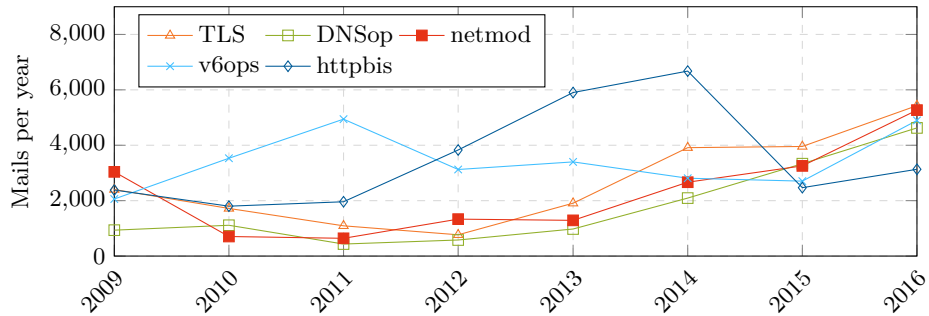


Fig. 1. Top 5 mailing lists of 2015 with most mails

Pure mailing list activity only constitutes one dimension of activity. Therefore, we also plotted the top 5 (of 2015) activity charts for counts of unique senders in Figure 2. The degrading of the netmod (from position 3 to 17), v6ops (4 to 11), and httpbis (5 to 8) mailing list are an artefact of the different types of discussions on the mailing lists: focused discussions that are conducted by a small group but in a quite interactive and responsive manner and statement-based discussions with different authors.

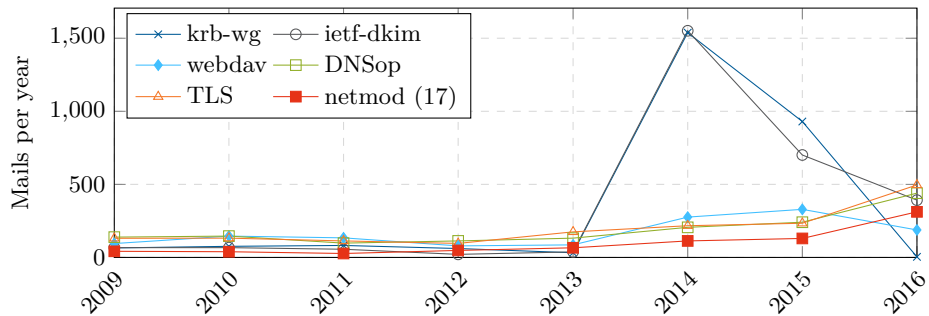


Fig. 2. Top 5 mailing lists of 2015 with most unique senders

The resulting analysis is based on the mails per sender as metric which we assume to be a naive but intuitive metric for the depth of discussions. Figure 3 shows the mailing lists with the highest average mails per sender in 2015.

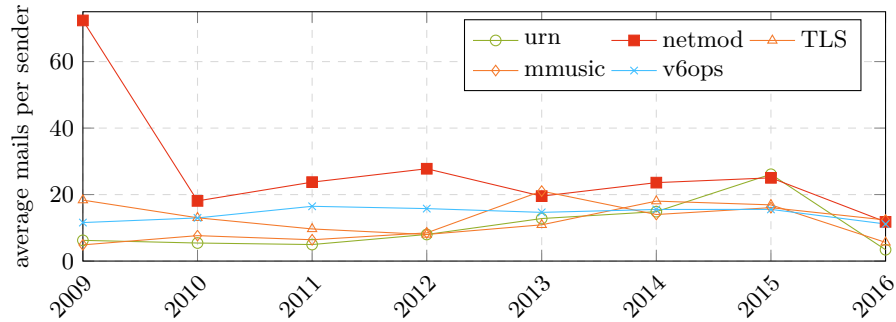


Fig. 3. Top 5 mailing lists with highest average mails per sender in 2015

3.2 Analysis of contributors to RFC documents

To conform with the open process in the IETF, the work done by individuals for the IETF is published and all discussion lists are open to the public. The IETF does not have official members. Individuals participating in the IETF do so on a voluntary basis. Therefore, a pay check affiliates contributors to a company or institution and entails a certain possibility of influence by the affiliated company or institution. Figure 4 shows the affiliation of document authors as stated in the RFCs. The interpretation of these authorship affiliations has to be made with care for two reasons: Specification of affiliation is voluntary according to the IETF and some persons may be affiliated twice, e.g. a professor affiliated to a company and a university. Similar statistics are generated by Jari Arkko [3].

Cisco is constantly over many years the company that contributes most to the writing of Internet Standards. In recent years Huawei has taken over the 2nd place, similar to its own rise to global player in the market. Now, the authors of actual standards tend to have such company affiliations and usually there are 2 to 4 such authors. If you look at the IETF mailing lists that discuss these standards, you will find much more activity. Figure 12(b) shows the number of distinct persons contributing to the TLS mailing list at the IETF. The TLS mailing list is currently (2014 - 2016) defining the new standard TLS 1.3 for secure communication over the Internet. The discussions on the TLS mailing list involves on average 60 and up to 100 distinct people each month. Here we see the amount of other volunteers who may have completely different affiliations. So, while important companies may seem to be relevant for the actual writing of a standard, many more volunteers from industry, academia, or independent contribute to how the standard solution will look like.

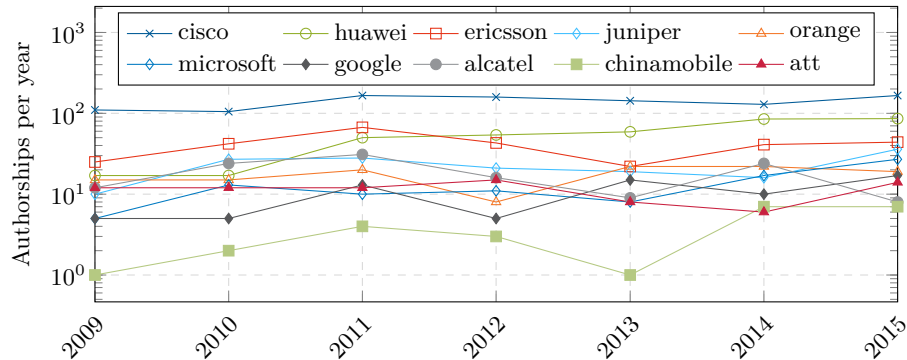


Fig. 4. Actors in the IETF standardisation

3.3 IETF-related activity on Twitter

Despite the official channels for discussion it is likely that discussions will also happen in side channels that are not directly made available by the IETF. Social networks are possible side channels for such discussions. Questions to be answered in that context include: Are there any IETF-related posts or tweets? Do they get liked or shared? Therefore, we utilized that IETF standards documents are called RFCs and they have a name and a number. E.g. RFC 791 is the original standard for the Internet Protocol (IP) from 1981. Recent standards have numbers between 7000 to 8000.

For our analysis we monitored the visibility of Internet Standards (IETF RFCs) on Twitter. We searched Twitter for tweets that contain the terms IETF, RFC-EDITOR, or RFC and a number, and then evaluate if it is an IETF-related tweet and if we can infer an RFC number from it. For the evaluation we use a combination of whitelisting (direct link to IETF or RFC-EDITOR sites) and blacklisting (common terms found in non-IETF tweets, e.g. from Football and Rugby clubs, user names referring to RFCs). Since Twitter only returns the most recent tweets for a search term, we regularly monitored Twitter to continuously log the most recent tweets since September 2015.

Figure 5 shows the activity of tweets related to Internet Standards in the time from September 2015 to February 2016. The average number of tweets per day is approximately 33. The peak early November is during IETF 94 meeting in Yokohama on November 1-6, 2015. No other IETF meeting occurred during the reported period.

Considering the size of Twitter and the large number of Internet Standards 33 tweets per day, do not seem a lot. If favourites and retweets are considered the activity doubles. There are on average 12 retweets and 18 favourites per day. Nonetheless, Twitter is not the center of IETF activity.

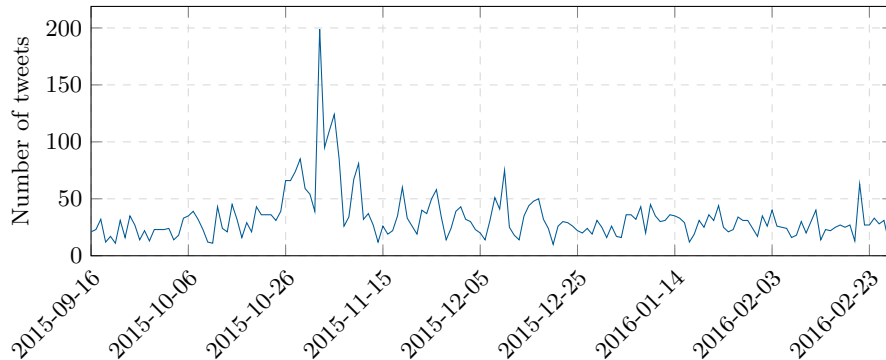


Fig. 5. Tweets per Day; between September 2015 and March 2016

3.4 RFC popularity on Twitter

Since we now know there is activity with respect to IETF standards on Twitter, we take a look at the standards that are discussed. Table 1 shows which Internet Standards got the most activity in February 2016. The first column shows the sum of tweets, retweets, and favourites.

Table 1. Twitter Statistics in February 2016

Activity	RFC	Name
109	6920	Naming Things with Hashes
103	822	STANDARD FOR THE FORMAT OF ARPA INTERNET TEXT MESSAGES
64	2324	Hyper Text Coffee Pot Control Protocol (HTCPCP/1.0)
58	3546	Transport Layer Security (TLS) Extensions
47	4204	Link Management Protocol (LMP)
44	7366	Encrypt-then-MAC for Transport Layer Security (TLS)
38	1149	Standard for the transmission of IP datagrams on avian carriers
36	7748	Elliptic Curves for Security
34	2549	IP over Avian Carriers with Quality of Service
34	2119	Key words for use in RFCs to Indicate Requirement Levels

A closer look at the tweets related to the RFC documents reveals the following relations: RFC 6920 is a proposed standard (not yet full Internet Standard) from April 2013. Most of the tweets simply name and link to the standard. Others additionally refer to the standard as being useful for their privacy protection. RFC 822 gets the tweets from people implementing related mail software. RFC 2324 is one of the famous April Fools RFCs, here a web protocol for the control of a coffee pot. The same is true for RFC 1149 and RFC 2549 on Internet over avian carriers like birds. RFC 3546 is present because of one tweet being favoured and

retweeted a lot. It embraces the 13 year old standard and refers to discussions (e.g. in the context of new TLS versions) to abandon the standard, which the tweet authors is an opponent of. RFC 4202 is present due to a Twitter bot that keeps tweeting about the standard and other arbitrary things regularly. The bot is a follow-back bot. RFC 7366 is a standard from late 2014, a single tweet asks about TLS implementations supporting the standard and rest is retweets and favourites. RFC 7748 is a recent standard from January 2016. RFC 2119 is a relevant old standard about word usage at IETF.

Similar listings can be produced for other periods of time in our measurement. In summary we see the following categories of RFC-related activity on Twitter:

- RECENT: promoting a comparably new document or standard (e.g. also for further discussion to make it full standard)
- RELEVANT: discussing an older RFC due to some current relevance
- FUN: sharing April Fools RFCs for fun (e.g. Hyper Text Coffee Pot Control Protocol (HTCPCP/1.0))
- BOT: a bot posting random tweets (one of which refers to RFC 4204). For the further analysis, we ignore the category BOT as it is only one bot posting older RFCs over and over again.

Socially-relevant RFCs like RFC 6920 on privacy-friendly naming and RFC 7686 on Tor’s .onion domain name (leading in October 2015 when the RFC was published) seem to be more visible in the top lists than arbitrary other recent RFCs.

In the following, we dive into the three main categories. For methodology: From 4-week-intervals ranging from October 2015 to end of February 2016, we took the top 10 most active RFCs of each month and further analysed their activity (tweets, retweets, favourite).

Figure 6 shows the events of all tweets from the RECENT category that made it into at least one 4-weekly top 10. 17 RFCs fall into this category. Dots indicate an event. The additional line indicates the months of publication of each RFC. As we do not have the data when tweets get retweeted and favoured, these are not contained in Figure 6.

A closer look at the activity data reveals: While some of the RFCs still show little activity after their initial peaks, these subsequent Twitter activities are very few and orders of magnitude lower than the initial peak. So, while the overall interest shows that these new RFCs are interesting to many people, there is yet no sign of viral activity. Compared to the initial peak, the activity afterwards declines to close to 0 in comparison. The large peaks, however, make it the category with most activity (2439 activities in the time interval, 143 per RFC).

Figures 7 and 8 show activities for the RELEVANT categories which includes older RFCs that receive some interest. In this case, only 7 RFCs made it into the list. In contrast to the tweets in the RECENT category that mostly announce RFCs, here there is usually a reference to an older RFC due to some relevance to a topic of interest, e.g. which IP addresses are private (RFC 1918 in the graph). The dots in Figure 7 show a much more continuous pattern over time than the related plot from the RECENT category. However, Figure 8 shows significant

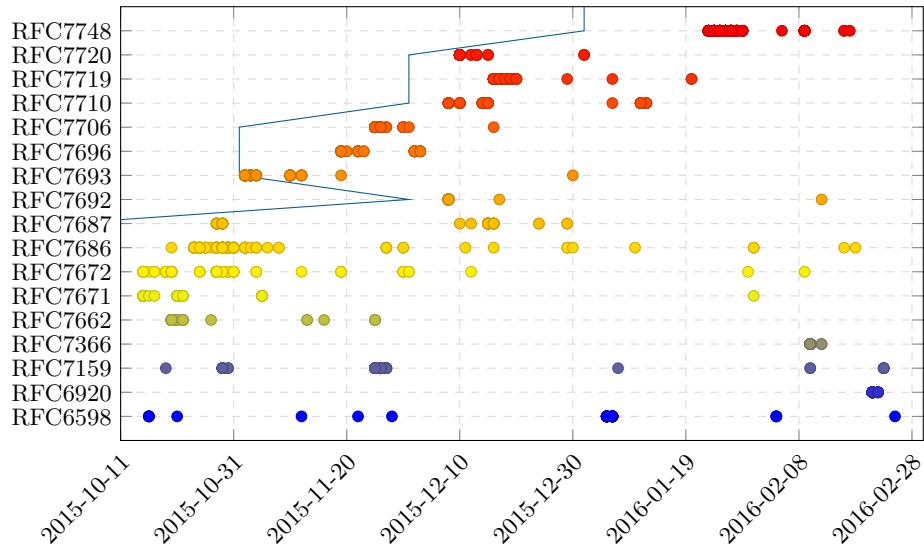


Fig. 6. Publication Date of RFC and Times when tweets about RFCs in category RECENT occurred

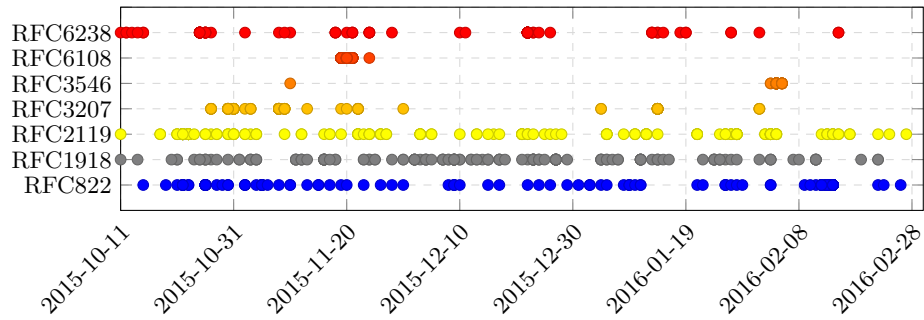


Fig. 7. Times when tweets about RFCs in category RELEVANT occurred

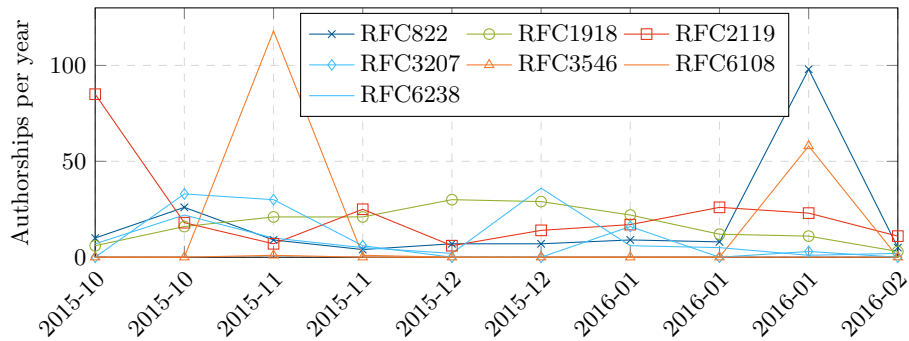


Fig. 8. Tweets per week in category RELEVANT

spikes of interest for some and more continuous interest for other RFCs. From the three categories, RELEVANT is the one that has the lowest activity (1119 activities in the time interval, 160 activities per RFC).

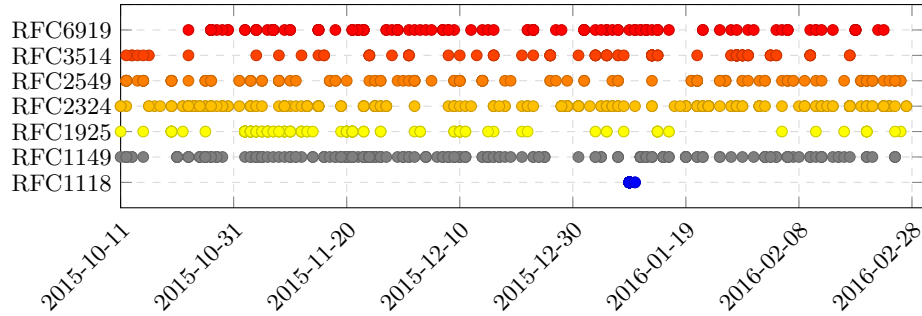


Fig. 9. Times when tweets about RFCs in category FUN occurred

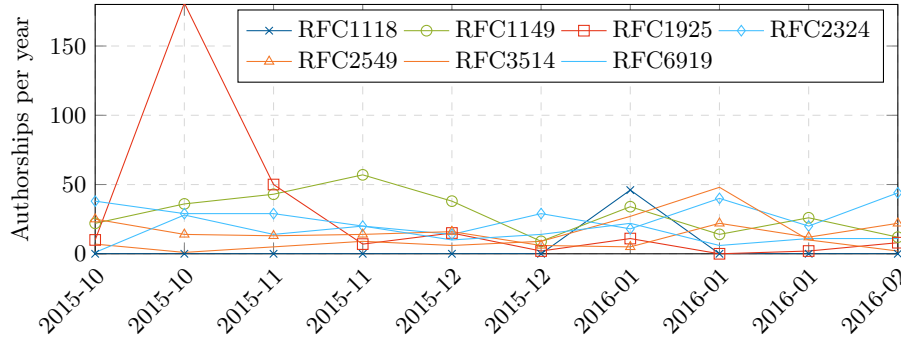


Fig. 10. Tweets per week in category FUN

Figures 9 and 10 are the related graphs for the FUN category that consists of standards published as April Fools RFCs on some April 1st. There are also 7 RFCs that made it to this list during the time we observed Twitter. Results displayed in Figure 9 follow a similar pattern as those in Figure 7. However, the graphs in Figure 10 reveal a more continuous interest into the individual RFCs than in the RELEVANT category. Here, both figures indicate a viral behaviour (for 6 of the 7 RFCs in the graph). From the three categories it has the 2nd highest activity (1314 activities in the time interval, 187 activities per RFC).

Figure 11 shows the amount of tweets, retweets, and favourites for each of the three categories. The RECENT category received a lot of retweets and favourites,

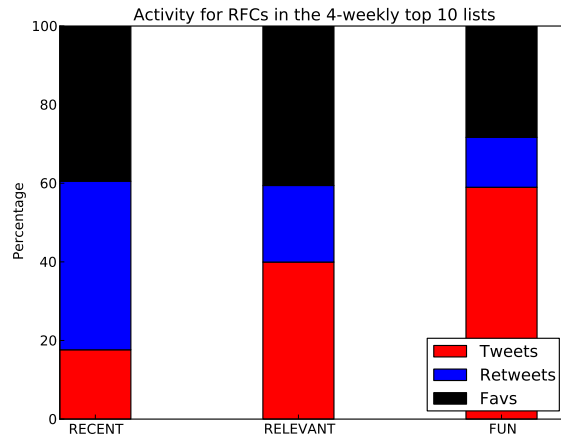


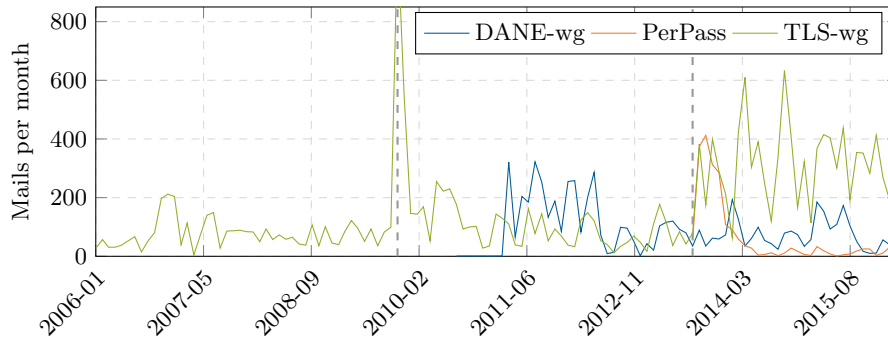
Fig. 11. Twitter activity type distribution concerning RFCs from the top 10 lists

but fewer actual tweets. The RELEVANT category has a similar percentage of favourites, but more actual tweets. The FUN category is dominated by tweets.

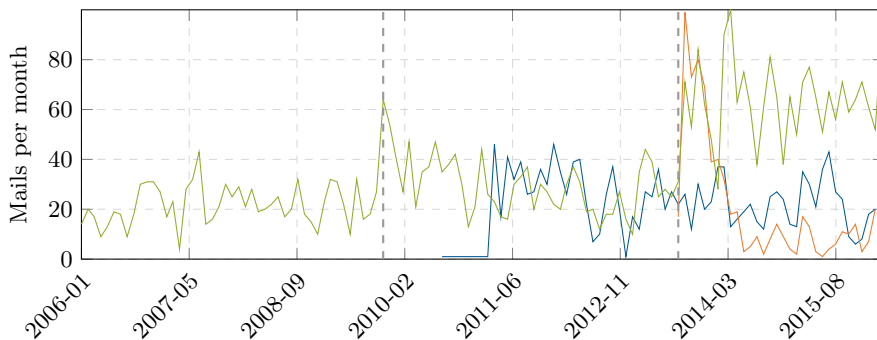
3.5 Impact of Snowden revelations on the IETF (Analysis of mailing lists)

On June 06, 2013 the Guardian and the Washington Post published the first articles about the NSA PRISM program. On June 09, 2013 the information was released that they obtained their information from Edward Snowden [9]. This brought the discussion of Internet security to a new height. Pervasive passive attackers have become a focus point of discussion. The IETF founded a new non-working group mailing list called PerPass on the subject of network and protocol design to mitigate pervasive monitoring [1].

Figure 12 shows the PerPass mailing list that was started a few weeks after the revelations and the security related working group mailing lists of DANE and TLS. We marked August 2013, the month of the Snowden leaks by dashed lines (right ones). As Figure12(a) shows, the number of mails immediately went into the orders of an important active mailing list like the TLS mailing list. In comparison the Snowden leaks have no direct impact on the DANE mailing list. Furthermore, Figure 12(b), below shows the number of different people that contributed each month. And indeed, a lot of different people contributed which shows the general interest into the subject. In the first half of 2014, however, the interest slowed down. There is still some activity, but it is minor compared to the initial peaks in late 2013. Nonetheless, the figure also shows increasing activities in the TLS working group that sets the standards for the most important security protocol of the web (TLS, which is, e.g. used as security in HTTPS).



(a) Total mails



(b) Mails with unique senders

Fig. 12. Mail activity on security-related mailing lists

When the previous version of TLS (version 1.2) was standardised in 2008, the mailing lists activities were only around 50 mails per month. Thus, the increase in interest cannot be explained with the standardisation of the next version (version 1.3). However, a similar peak with even up to 1097 mails per month (November 2009, left dashed line in Figure 12) was reached for a short period of time end of 2009 when a severe security flaw was found in TLS which was fixed with RFC 5746 in February 2010. Interestingly, the number of distinct senders (different people) that contributed on the mailing list in the interval only peaked at 64 different senders, which is roughly the number it currently converged to as normal number of distinct contributors. The all-time monthly peak was 100 in April 2014.

The IETF had to react to the Snowden revelations. However, concluding from the statistics of our observation, we can see that IETF reaction to the revelations of Edward Snowden was indeed profound and spawned a lot of interest and participation in related IETF activities. Using the TLS mailing list again, Figure 12 shows that its activity before and around 2008 was much smaller, even though the previous TLS standard was produced back then.

4 Conclusions and Outlook

The authors of IETF standards are usually associated with an Internet-related tech company, Cisco and Huawei being currently the largest contributors. Looking only at the actual authors is, however, misleading as a much larger amount of people takes part in the discussions of the standardisation process.

With respect to social media, interest in IETF standards (RFCs) exists, yet it is not very high. We classified the Twitter activities in tweets about recent RFCs, about older relevant RFCs, and funny April Fools RFCs. The latter are the most viral.

Mailing list activities show that societal influences like the Snowden revelations as well as attacks found against a security standard both trigger increased activities. For Snowden, the first increase went to a freshly established mailing list on the issue, then it progressed to other security activities. The data suggests that it most likely influenced and increased IETF participation.

The presented graphs are extracted from our continuously updating IETF activity database. In future we plan to provide web access to the dataset and to include further data sources beside the IETF web pages, mailing lists, and Twitter.

References

1. Perpass – the perpass list is for ietf discussion of pervasive monitoring. <https://www.ietf.org/mailman/listinfo/perpass>. [last checked: 1.4.2016].
2. H. Alvestrand. A Mission Statement for the IETF. RFC 3935 (Best Current Practice), October 2004.
3. Jari Arkko. IETF Statistics: Fun and Useful Trivia. "<http://www.arkko.com/tools/docstats.html>". [last checked: 1.4.2016].
4. S. Bradner. IETF Working Group Guidelines and Procedures. RFC 2418 (Best Current Practice), September 1998. Updated by RFCs 3934, 7475.
5. Scott Bradner. IETF Structure and Internet Standards Process. Presentation at the 81st IETF, 2011.
6. Vint Cerf. IETF and the Internet Society. Webpage, 1995.
7. Sam Ockman Chris DiBona. *Open Sources: Voices from the Open Source Revolution*. O'Reilly Media, 1999.
8. D. Crocker and N. Clark. An IETF with Much Diversity and Professional Conduct. RFC 7704 (Informational), November 2015.
9. Mirren Gidda. Edward snowden and the nsa files timeline. <http://www.theguardian.com/world/2013/jun/23/edward-snowden-nsa-files-timeline>, August 2013. [last checked: 1.4.2016].
10. P. Hoffman and S. Harris. The Tao of IETF - A Novice's Guide to the Internet Engineering Task Force. Webpage as specified in RFC 6722, 2012.
11. R. Hovey and S. Bradner. The Organizations Involved in the IETF Standards Process. RFC 2028 (Best Current Practice), October 1996. Updated by RFCs 3668, 3979.
12. M. Kucherawy. IAB, IESG, and IAOC Selection, Confirmation, and Recall Process: Operation of the Nominating and Recall Committees. RFC 7437 (Best Current Practice), January 2015.

13. The Apache SpamAssassin Project. SpamAssassin. <http://spamassassin.apache.org>. [last checked: 1.4.2016].
14. The Apache SpamAssassin Project. Tests Performed: v3.3.x. http://spamassassin.apache.org/tests_3_3_x.html.