A Pragmatic Approach for Reducing Delays in Publishing Documents within the Real-time Applications and Infrastructure (RAI) Area
draft-tschofenig-rai-reducing-delays-01.txt

Status of this Memo

Copyright Notice

Abstract

During the last year, participants in the Real-time Applications and Infrastructure (RAI) area have been quite active in discussing

proposals that could improve their way of working. This document is a contribution to that discussion and focuses on the reduction of delays experienced in producing specifications. We believe that this is one of the main problems in the RAI area (and quite likely in other areas of the IETF as well) and it requires attention. A number of side effects, caused by the long specification work, are illustrated in this document.

Table of Contents

---

1. **Introduction**   TOC

Many participants in the Real-time Applications and Infrastructure (RAI) area have noticed that it takes several years for a specification from the initial working group item to the published RFC. A brief look at http://www.arkko.com/tools/lifecycle/ (for documents like ICE, SIP Location Conveyance, SIP configuration framework) very quickly confirms this observation. Several years are quickly spent. In various cases these delays have had negative effects on the deployment situation and on interoperability. In some other cases the relationship between a long delay and lack of deployment cannot be directly observed and there are more complex reasons that can only be analysed on a per-specification basis. For example, some of the interoperability problems that got discussed during the SIP Interoperability Workshop at IETF#70 (see http://www.sipforum.org/component/option,com_docman/task,cat_view/gid,42/Itemid,75) may have been caused by specifications with too many options.
The authors are not aware of a detailed and structured analysis of the documents within the RAI area with respect to their delays. With some documents there is the question whether additional delay is actually problematic, particularly when they have a research character. One could argue that optimizations to the process should not be started without a detailed analysis of the main causes for these delays. Unfortunately, the delays are the effects of actions taken by individuals and blaming them for the problems is not helpful, particularly since the accused persons are very likely to disagree.

Protocol specifications that are not deployed are useless regardless
how fast they found their way through the IETF. This document, however,
does not attempt to discuss problems of publishing specifications that
do not meet market demands (especially since different IETF
participants target different market segments, some of which might not
even closely reflect the properties of the Internet).
Instead, this document only focuss on delays assuming that the only
problem to be solved is that specifications are not published in a
timely fashion.

---

2. Reasons for Delays

The authors believe that the following list provides a fairly complete
list of reasons for delays in the process of publishing a
specification:

Interworking between different working groups:
While it may sound
fairly unproblematic to work on a single specification in
different working groups experience has shown the opposite.
Typically, the problems are related to the lack of understanding
of the usage scenarios, different schedules and priorities,
unclear responsibilities, different workstyles in different
groups, etc. A similar problem can also be observed when special
expert groups need to get involved, e.g., in the form of XML
directorates, reviews of URI schemes, application layer design
aspects.

Examples that support the above statement can be found with
RADIUS GEOPRIV [I-D.ietf-geopriv-radius-lo] (Tschofenig, H.,
Adrangi, F., Jones, M., Lior, A., and B. Aboba, "Carrying
Location Objects in RADIUS and Diameter," May 2009.) (interaction
between GEOPRIV and RADEXT), HELD
[I-D.ietf-geopriv-http-location-delivery] (Barnes, M.,
Winterbottom, J., Thomson, M., and B. Stark, "HTTP Enabled
Location Delivery (HELD)," August 2009.) (interaction between
GEOPRIV and the APPS area), and SIP Location Conveyance
[I-D.ietf-sip-location-conveyance] (Polk, J. and B. Rosen,
"Location Conveyance for the Session Initiation Protocol,"
March 2009.) (interaction between GEOPRIV and SIP).

Interworking with other SDOs:
Similarly to the interworking between
different IETF working groups one might already expect problems
in the interworking with other SDOs as the differences might
quickly increase due to the different set of participants, fairly
different standardization procedures and deadlines ('yes, in some

SDOs deadlines are taken serious even though they are completely
artificial'), different architectural approaches, different
business models, different target markets (e.g., enterprise
networks, cellular networks, military networks) different
security assumptions.

Examples can be found throughout the IETF. In the RAI area the
SIP change process (see [RFC3427] (Mankin, A., Bradner, S., Mahy,
R., Willis, D., Ott, J., and B. Rosen, "Change Process for the
Session Initiation Protocol (SIP)," December 2002.) and the
currently discussed revision of it [I-D.peterson-rai-rfc3427bis]
(Peterson, J., Jennings, C., and R. Sparks, "Change Process for
the Session Initiation Protocol (SIP) and the Real- time
Applications and Infrastructure Area," October 2009.)) build the
basis of the interworking. Unfortunately, it turned out that in
practice problems show up with the lack of participation of
persons in both organizations. A high priority item from, for
example, the TISPAN leads to a lot of confusion and lack of
interest in the IETF RAI area due to the lack of context. There
are, however, different models in the IETF as well where a lot of
the actual protocol work is outsourced to the respective
organization. An example would be the RADEXT working group with
their work on RADIUS [RFC2865] (Rigney, C., Willens, S., Rubens,
A., and W. Simpson, "Remote Authentication Dial In User Service
(RADIUS)," June 2000.); a protocol that is being used by a number
of SDOs. Although the specifications that are sometimes produced
outside the IETF do not meet the quality expectations in the IETF
they do at least not cause resources from IETF participants to be
wasted when there is little to no interest in that specific work.
To provide a minimum degree of guidance rules have been outlined
that should be followed when defining new extensions for RADIUS,
see [I-D.ietf-radext-design] (Weber, G. and A. DeKok, "RADIUS
Design Guidelines," April 2010.).

Disagreement about architectural approaches:

Sometimes document
authors have a different architectural approach in mind that is
incompatible with a large part of the working group members.
Often, these aspects are discovered fairly late in the process
particularly when the document lacks a description of the
assumptions and the security architecture. Often, the problems
are related to the interworking with other SDOs as the solution
approach developed with the IETF specification as seen by the
authors as a building block that has then to be fit into an
architecture developed elsewhere (often without explicitly
stating who else would be doing that work).

Examples include the stream of work regarding GETS/MLPP
(partially done in the IEPREP working group).

Review marathon:

Getting a document through the IETF process does not happen
without reviews. There are reviews before the document becomes a
WG item, reviews while the document is within the group, some
chairs demand reviews before they issue a WGLC to probe
'readiness', someone in between reviews by special expert groups
are done (also by other SDOs and external groups, if necessary),
then comes the WGLC, review by the responsible AD, review by the
IESG, review by directorates (and there are many of those) and
(depending on the type of document and history) an IETF Last
Call. Everybody has an opinion, often not necessarily a technical
opinion, on how documents should be written and why other
solution approaches have not been explored. Reviewers need time
and then the review comments often cannot be ignored but need to
be discussed and resolved. When reviews happen later in the
process then text changes are often expected to keep the reviewer
happy. IESG members frequently put DISCUSSes on reviews and this
increases their priority allowing a single person to, for
example, delay the publication of a document for an extended
period of time. From a psychological point of view reviewers are
in the unfortunate position that they have the feeling that
something must be improved as an outcome of the review activity.
As soon as documents leave the working group the transparency is
largely lost, despite IESG comments being sent to the authors, WG
chairs and responsible ADs and despite information being
available in the I-D tracker. Mentally, many working group
members consider documents to be 'done' when they leave the
working group.

The authors are not arguing that reviews are unnecessary but
there has to be balance with respect to the goal that is about to
be accomplished.

Degeneration of the 3-level Standards Track process:
                                                      RFC 2026
[RFC2026] (Bradner, S., "The Internet Standards Process --
Revision 3," October 1996.) describes the 'Internet Standards
Process' and describes the level of Standards Track documents,
namely 'Proposed Standard' to 'Draft Standard' to 'Internet
Standard'. It appears that IETF participants are less likely to
spend their time on advancing documents from 'Proposed Standard'
to 'Draft Standard', particularly since the industry to a large
extent does not understand the standards process anyway. More and
more it can be observed how a fairly small group of people get
together and produce de-facto standards by finishing
specifications in an astonishing time frame together with a large
number of implementations that the Internet community is willing
to deploy. The standards process utilized in the IETF is,

unfortunately, not tailored to these type of developments. Experimental documents on the other hand, although easier to publish through the IETF, have the stigma of being very unbaked. In certain environments the label of 'experimental' is considered as a problem. Putting normative references to informational or experimental RFCs in Standards Track documents later makes the publication process more complex due to the need for Down-Refs [RFC3967] (Bush, R. and T. Narten, "Clarifying when Standards Track Documents may Refer Normatively to Documents at a Lower Level," December 2004.).

Feature creep and overall complexity:
Setting up a working group takes some time, getting working group members to agree that a certain document has to become WG item also takes a lot of time, the time it takes to publish it as an RFC takes even longer and producing more extensions involves also a certain overhead (e.g., re-chartering, new working group/BOF, new milestones, support from the working group). It is therefore not surprising that document authors/edtiors or the working group are sometimes tempted to add a tiny feature here and another small feature there.

A further favorite 'passion' is to generalize the solution. In theory this is a good idea since the same protocol can then be used in a variety of different usage scenarios. In most case, this has lead to more complexity, much longer time to finish the specification and sometimes none of the use cases are consequently covered appropriately. For some, the SIP configuration framework might be an example.

Lack of time:
The IETF to a large extend consists of volunteers (rather than standardization specialists and consultants), who have a fairly limited time budget. Those who are involved in the development of standards tend to have time constraints (e.g., authors, WG participants, WG chairs, ADs, Expert Reviewers, etc.). This is quite normal.

Problems arise when the lack of time causes considerable delays in completing work other people rely on. Consider the following hypothetical case. Bob is an expert in writing specifications and he has a lot of energy and enthusiasm. He is assigned as the main editor of a specification. The work on the specification takes some time (typically years) but Bob does his job well and the community respects him. Over time Bob develops new interests, might even change his job and does not show the same availability anymore. Still, Bob remains in charge of the document and Bob does not recognize that he is unable to commit the necessary time

to get the job done. A similar situation can happen in any other role previously mentioned.

Missing running code:
Although most people would agree that running code is very useful many specifications are not backed up with it. Since specifications change a lot over the lifetime before being published as an RFC (even at a very late stage) it can be pretty frustrating to have running code in an early stage. The standards process considers interoperable code but due to the long time it takes to publish specifications it is rarely utilized.

As an example, the TURN specification [I-D.ietf-behave-turn] (Rosenberg, J., Mahy, R., and P. Matthews, "Traversal Using Relays around NAT (TURN): Relay Extensions to Session Traversal Utilities for NAT (STUN)," July 2009.) has changed quite considerably over time. Those who implemented earlier versions essentially had to re-write their code.

Onerous extensibility rules:
Many document authors feel insecure when writing IANA consideration sections and for some reasons these consideration sections often demand Standards Track documents to introduce new extensions.

A recent example can be found with the Service URN document [RFC5031] (Schulzrinne, H., "A Uniform Resource Name (URN) for Emergency and Other Well-Known Services," January 2008.), which requires a Standards Track document for allocating a new top-level top-level service labels. With the new extensions that the working group had seen it became cumbersome to progress these documents in a timely fashion through the working group.

AD and IESG overload:
Some documents need significant time after leaving the working group until they are published. It is unclear that these delays yield significantly better documents in the end. Part of the problem is that the current IETF area director model seems to assume that ADs spend the equivalent of a full-time job on their "volunteer" job, which appears to be unrealistic, particularly when ADs have been in office for several years. Compared to other organizations, the AD "span of control" is very large, supervising twenty or more working group chairs.

Exhaustion:
As the publication process wears on, the amount of change per time unit continuously decreases. Because of the long delays, authors are often working on many documents in parallel,

making it difficult for them to switch context and remember all
the issues when they update documents. As the initial excitement
about a problem fades after a few years, editing such documents
becomes a chore, further delaying the publication. This leads to
author exhaustion.

Thus, to the extent possible, most documents should be finished
in no more than three IETF meetings: gauge interest and agreement
on basic approach during the first IETF meeting and assign
reviewers, reach consensus on the technical issues during the
second IETF meeting and get a final WG approval during the third
one, if needed.

---

3. Suggestions

Below, we present a set of suggestions to reduce publication delays. We
believe that many of the suggestions can be used together, and some can
be used experimentally to see if they produce significant improvements.

Improving the chartering process:
It takes far too long to add an
item to a charter. If there's interest in the working group and
the document is below a certain size and complexity (based on the
judgement by the WG chair) and three reviewer volunteers, the
document should just get into the next cycle.

State assumptions clearly:
Sometimes a specific protocol
specification is target for usage in a certain environment (or
envisioned context). The working group may be fine with such an
approach but often the document does not state the usage
environment. This often leads to confusion in the review process.
For example, some documents are being progressed through the IETF
where the main use case can be found in the 3GPP IMS
architecture. Maybe the document assumes that other organizations
utilize the specification in a specific context. There are often
additional operational aspects that need to be considered in
order to come to a complete solution.

A recent example can be found in the work on
[I-D.ietf-ecrit-local-emergency-rph-namespace] (Polk, J., "IANA
Registering a SIP Resource Priority Header Field Namespace for
Local Emergency Communications," March 2010.) where the document
essentially only specifies the syntax but the semantic is
supposed to be provided by organizations outside the IETF.

Delegate down:

In most organizations, middle management is given significant decision authority, usually described by the financial impact of a decision. A CEO of company does not approve every contract, user manual or feature list. The current IETF review model is predicated on the notion that the IESG reviews protocols to ensure that they do not damage the Internet. However, most of the current RAI documents are extensions and bug fixes that are likely to have only local effects.

To reduce AD and IESG load, working group chairs should be able to handle most documents until they reach the RFC editor, unless there is an objection during IETF last call. IESG members are expected to voice their objection during that last call period, triggering formal IESG review. Only efforts that introduce new protocols or are likely to have significant Internet-wide impact require more extensive review.

When a document is added to the WG work list, it should be flagged to indicate that it will require a full review. All other documents receive a lighter late-stage review.

Clearly label IETF discussion slots:
Given the three-meeting model mentioned earlier, the IETF meeting presentations should be structured accordingly, with a clear set of questions to be answered. For example, the first presentation needs to answer questions about architectural assumptions, relation to other work, dependencies, importance of the problem and alternative approaches. Working group chairs should work with document authors to structure presentations, possibly providing templates. In this model, the second discussion is crucial and should be given ample time, if necessary enhancing the plenary working group meeting with break-out meetings before and after the WG meeting slot. (It is a waste of time to have 200 people listen to a discussion where only a handful can really make significant contributions.)

Journal-like review model:
When a new document is submitted to the working group, the working group chair determines at latest at the next IETF meeting whether there is sufficient interest to pursue this work. If so, he or she seeks at least three reviewers from the working group. If such reviewers cannot be found, the document lacks sufficient interest. Authors may suggest qualified reviewers. Reviewers are given a two-month review deadline, so that reviews would typically arrive mid-cycle between IETF meetings. The basic process could follow the GenArt review model, which could proceed in parallel. As needed, a specialized

reviewer might be assigned to only look at specific issues, such as XML usage, security aspects or congestion control.

The reviews should be tracked in an issue tracker, in addition to being disseminated via the working group mailing list. As a short-term measure until the IETF tools are enhanced, existing conference review tools might be used, as they offer definable review forms (e.g., to separate editorial and technical issues) and automated review reminder mechanisms.

This approach closely models the review process used by scientific journals and technical conferences.

Increasing implementation feedback:
When documents are able to progress faster through the IETF participants should feel encouraged to provide an update that includes bug fixes and other corrections based on implementation efforts. A possible approach to accomplish this today is to publish documents as informational and experimental RFCs (faster) and to advance them to Standars Track once implementation (or even deployment) experience is available.

Delegate work to other SDOs:
We believe that the updated version of the SIP Change Process [I-D.peterson-rai-rfc3427bis] (Peterson, J., Jennings, C., and R. Sparks, "Change Process for the Session Initiation Protocol (SIP) and the Real- time Applications and Infrastructure Area," October 2009.) goes into the right direction already.

Additional WG chair training:
We suggest to use some of the WG chair training lunches to offer room for discussions on how to avoid delays in progressing documents and to collect the best current practises.

Virtual interim meetings:
Many document authors do their IETF work in the two-week period around the submission deadlines before an IETF meeting. This leads to fairly low document update cycles. A tool to increase the time that is spent on documents per IETF cycle working style is to hold virtual interim meetings (i.e., phone conferences). This is common practice also in other organizations and helps to stay focused on open issues. Sometimes chairs regularly contact the main document authors to discuss the editing progress and the status of open issues. It is useful to allow working group members to participate and distribute notes about this informal communication.

Keep track of the WG status:

Sometimes no progress is made because there is uncertainty about who is responsible for taking the next step. Periodic reminders (even with tool support) would help to make the workflow more visible. Examples of these periodic status updates can be found here http://www.ietf.org/mail-archive/web/ecrit/current/ msg05785.html, http://www.ietf.org/mail-archive/web/ops-area/ current/msg00467.html and http://www.ietf.org/mail-archive/web/ saag/current/msg02242.html.

Document status updates:
Each document must have an dedicated editor. This person is responsible for making sure that the status of the document is known. Ensuring a regular interval for reporting is helpful, such as once a month. The fact that such a short report (a few lines in most cases) is written already reminds the editor to look at open issues, which often helps to make progress.

Too often, document editors erroneously believe that it is the responsibility of the working group to drive the process and the lack of feedback is an indication that the document is ready.

Keep your ADs informed:
As previously argued it is important to keep the working group informed about the current status. In the same style it is important to let the ADs know what the current status of the work is. Regular phone conference calls, e.g. once a month, are a possible approach.

Assign new editors:
If the original author clearly cannot complete the work in a timely fashion, the working group chair should routinely re-assign an editor to the document, after, say, a one-year delay since the -00 draft. If no other person is capable to take that role, this indicates that the document is either too long, too complicated or of insufficient interest to merit further consideration as-is and needs to be split up, simplified or abandoned.

Enhance I-D tracker:
The I-D tracker should indicate who is supposed to be reviewing the document and by what deadline. It should also clearly indicate who has the "token" on the document, i.e, whether the next action is expected to come from the working

group chair, a reviewer or set of reviewers, a directorate, or
the authors.

Posting minutes:
                    The working group chair(s) must publish the
meeting minutes (of a face-to-face or virtual interim meeting)
within a few days after the meeting. This helps to ensure that
action items can be captured by the participants. Too often
details are lost with longer delays and the benefit of the
meeting may get lost.

Prepare meeting slides appropriately:
                                        Slides are a common way to
present open issues and to provide a status update during
(virtual) interm and face-to-face meetings. Often the slides are
available to working group chairs a few minutes before the
meeting starts. Furthermore, the style of slides often does not
serve as a suitable form to discuss the content. Since small time
slots are allocated there is no time to get many of the
participants up-to-speed with the discussed issue leaving a large
number of the participants clueless. Although it is the seen as
the responsibility of the participant to read into the subject it
is also questionable whether a presentation with only a handfull
of persons understanding the issue is really helpful. Instead,
the document editor and the document authors should consider
using alternative means of preparing the content in a suitable
format to ensure that a large number of persons are actually able
to provide a feedback. An example, for such an approach can be
found with blog posts http://www.hueniverse.com/hueniverse/
2009/04/explaining-the-oauth-session-fixation-attack.html or
http://www.hueniverse.com/hueniverse/2008/10/beginners-
gui-1.html.

The IETF meetings are about discussion and discussions need to be
prepared. Not everybody is able to have get a definitive opinion
while listening to a 5-10 minute (particularly when the topic is
complex). Consider that not everyone is given with the gift to
give presentations that are easy to understand (considering also
the language barrier). Consequently, if the audience is not
prepared then there is no discussion. Trying to close open issues
will be difficult.

Stay connected:
                Exchanging information about ongoing activies and
challenges in a timely fashion it is vital to have other persons
contact information (e.g. IM nicknames, various phone numbers,
etc.). An email conversation is often not timely enough. A
working group chair, for example, should at least have the

> contact details of the ADs responsible for the working group, the
> document authors and the key working group members.

---

4. Security Considerations

This document discusses process related aspects that are not directly
related to security. However, the outcome of an improved process might
have a positive impact on security provided by the deployed
specifications.

---

5. Acknowledgements

We would like to thank all of those you care about the process related
aspects in the IETF as they contribute to the impact of the work
produced by the IETF in the future. We would also like to thank
everyone participating on the RAI area mailing list for their input
during the RAI reorganization discussions.
We would particularly like to thank Benoit Claise for his feedback to
the document.

---

6. Informative References

[I-D.ietf-behave-turn]	Rosenberg, J., Mahy, R., and P. Matthews, "Traversal Using Relays around NAT (TURN): Relay Extensions to Session Traversal Utilities for NAT (STUN)," draft-ietf-behave-turn-16 (work in progress), July 2009 (TXT).
[I-D.ietf-ecrit-local-emergency-rph-namespace]	Polk, J., "IANA Registering a SIP Resource Priority Header Field Namespace for Local Emergency Communications," draft-ietf-ecrit-local-emergency-rph-namespace-04 (work in progress), March 2010 (TXT).
[I-D.ietf-geopriv-http-location-delivery]	Barnes, M., Winterbottom, J., Thomson, M., and B. Stark, "HTTP Enabled Location Delivery (HELD)," draft-ietf-geopriv-http-location-delivery-16 (work in progress), August 2009 (TXT).
[I-D.ietf-geopriv-radius-lo]	Tschofenig, H., Adrangi, F., Jones, M., Lior, A., and B. Aboba, "Carrying Location Objects in RADIUS and Diameter," draft-ietf-geopriv-radius-lo-24 (work in progress), May 2009 (TXT).

[I-D.ietf-radext-design]	Weber, G. and A. DeKok, "RADIUS Design Guidelines," draft-ietf-radext-design-13 (work in progress), April 2010 (TXT).
[I-D.ietf-sip-location-conveyance]	Polk, J. and B. Rosen, "Location Conveyance for the Session Initiation Protocol," draft-ietf-sip-location-conveyance-13 (work in progress), March 2009 (TXT).
[I-D.peterson-rai-rfc3427bis]	Peterson, J., Jennings, C., and R. Sparks, "Change Process for the Session Initiation Protocol (SIP) and the Real- time Applications and Infrastructure Area," draft-peterson-rai-rfc3427bis-04 (work in progress), October 2009 (TXT).
[RFC2026]	Bradner, S., "The Internet Standards Process -- Revision 3," BCP 9, RFC 2026, October 1996 (TXT).
[RFC2865]	Rigney, C., Willens, S., Rubens, A., and W. Simpson, "Remote Authentication Dial In User Service (RADIUS)," RFC 2865, June 2000 (TXT).
[RFC3427]	Mankin, A., Bradner, S., Mahy, R., Willis, D., Ott, J., and B. Rosen, "Change Process for the Session Initiation Protocol (SIP)," RFC 3427, December 2002 (TXT).
[RFC3967]	Bush, R. and T. Narten, "Clarifying when Standards Track Documents may Refer Normatively to Documents at a Lower Level," BCP 97, RFC 3967, December 2004 (TXT).
[RFC5031]	Schulzrinne, H., "A Uniform Resource Name (URN) for Emergency and Other Well-Known Services," RFC 5031, January 2008 (TXT).

Authors' Addresses

TOC

	Hannes Tschofenig
	Nokia Siemens Networks
	Linnoitustie 6
	Espoo 02600
	Finland
Phone:	+358 (50) 4871445
Email:	Hannes.Tschofenig@gmx.net
URI:	http://www.tschofenig.priv.at
	Henning Schulzrinne
	Columbia University
	Department of Computer Science
	450 Computer Science Building
	New York, NY 10027
	US

Phone:	+1 212 939 7004
Email:	hgs+ecrit@cs.columbia.edu
URI:	http://www.cs.columbia.edu
	Markus Isomaki
	Nokia
	P.O.BOX 100
	00045 NOKIA GROUP
	Helsinki
	Finland
Phone:	
Email:	markus.isomaki@nokia.com