Abstract

The Session Initiation Protocol (SIP) "Caller Preferences" extension defined in RFC 3840 provides a mechanism that allows a SIP message to convey information relating to the originator’s capabilities. This document makes it possible for SIP proxies to convey similar information, by extending the rr-param rule defined in RFC 3261, so that the header field parameter can be used to convey feature tags that indicate features supported by the proxy.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on March 28, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of
the Trust Legal Provisions and are provided without warranty as
described in the Simplified BSD License.

Table of Contents

1. Introduction ........................................... 3
   1.1. Use-case: IMS Service Continuity .................. 3
2. Conventions ........................................... 3
3. Definitions ........................................... 3
4. User Agent behavior ................................... 4
5. Proxy behavior ........................................ 4
6. Feature tag semantics .................................. 4
7. Examples .............................................. 5
   7.1. Example name ...................................... 5
8. IANA Considerations .................................... 5
9. Security Considerations ................................ 5
10. Acknowledgements ..................................... 5
11. Change Log ........................................... 5
12. References ........................................... 6
   12.1. Normative References .............................. 6
   12.2. Informative References ............................ 6
Authors’ Addresses ......................................... 6
1. Introduction

The SIP "Caller Preferences" extension defined in RFC 3840 [RFC3840] provides a mechanism that allows a SIP message to convey information, using feature tags, relating to the originator's capabilities.

Feature information can be useful for other SIP entities, that might trigger actions and enable functions based on features supported by other SIP entities.

This document extends the rr-param rule defined in RFC 3261 [RFC3261], so that it can be used to convey feature tags indicating support of features in SIP proxies. The rr-param rule is used in the SIP Path, Route, Record-Route and Service-Route header fields.

1.1. Use-case: IMS Service Continuity

The 3rd Generation Partnership Project (3GPP) defines a IP Multimedia Subsystem (IMS) Service Continuity mechanism [3GPP.23.237] for handover of Packet Switched (PS) sessions to Circuit Switched (CS). The handover can be performed by a Service Centralization and Continuity Application Server (SCC AS), or by a SCC AS together with an Access Transfer Control Function (ATCF), that acts as a SIP proxy. Delegating part of the session handover functionality to an ATCF provides advantages related to voice interruption during session handover etc, since it is located in the same network as the user.

In order for a SCC AS to delegate part of the session handover functionality to an ATCF, when it receives a SIP REGISTER request, it needs to be informed whether there is a proxy that provides ATCF functionality in the registration path.

2. Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14, RFC 2119 [RFC2119].

3. Definitions

The rr-param rule defined in RFC 3261 [RFC3261]:

        rr-param = generic-param

is extended to:

Holmberg & Sedlacek   Expires March 28, 2011
Internet-Draft  proxy feature  September 2010

rr-param = generic-param / feature-param

where feature-param is defined in Section 9 of RFC 3840 [RFC3840].

4. User Agent behavior

This specification does not specify any new User Agent behavior.

5. Proxy behavior

When a proxy inserts a Path header field (during registration), a Service-Route header field (during registration) or a Record-Route header field (during a dialog establishment), it MAY insert a feature tag in the header field.

If a feature tag is inserted in a Path or Service-Route header field during registration, the resource identified by the URI in the header field MUST provide support for the associated feature for all dialogs associated with the registration, until the registration is terminated or re-freshed.

If a feature tag is inserted in a Record-Route header field during a dialog establishment, the resource identified by the URI in the header field MUST provide support for the associated feature until the dialog is terminated.

6. Feature tag semantics

The feature tag in a header field constructed using rr-param rule indicates support of the feature in the resource identified by the URI in the header field.

In order to insert a feature tag in a SIP header field constructed by using rr-param rule, the feature specification MUST specify the semantics of the feature tag when inserted in that specific header field. Unless the feature specification defines such semantics, the feature tag MUST NOT be included in that specific header field.

NOTE: If a route set is built using Path, Record-Route or Service-Route header fields, any inserted feature tag will be copied into the associated Route header fields, together with other header field parameters. This specification does not define any specific meaning of the feature tags present in Route header fields in such cases.
7. Examples

7.1. Example name

TBD

\begin{center}
\begin{tabular}{c|c|c}
Alice & P1 & REGISTRAR \\
\hline
--- REGISTER----------> & --- REGISTER---------->
 & Path: P1;+g.3gpp.srvcc \\
\hline
 & <-- 200 OK ----------
 & Path: P1;+g.3gpp.srvcc
 & Service-Route: REG \\
\hline
 <-- 200 OK ----------
 & Path: P1;+g.3gpp.srvcc
 & Service-Route: REG \\
\hline
\end{tabular}
\end{center}

Figure 1: Example call flow

8. IANA Considerations

TBD

9. Security Considerations

Feature tags can provide sensitive information about a SIP entity. RFC 3840 cautions against providing sensitive information to another party. Once this information is given out, any use may be made of it.

10. Acknowledgements

Thanks to Paul Kyzivat for his comments and guidance on the mailing list.

11. Change Log

[RFCEditor NOTE: Please remove this section when publishing]
12. References

12.1. Normative References


12.2. Informative References


Authors’ Addresses

Christer Holmberg
Ericsson
Hirsalantie 11
Jorvas 02420
Finland

Email: christer.holmberg@ericsson.com

Ivo Sedlacek
Ericsson
Scheelevaegen 19C
Lund 22363
Sweden

Email: ivo.sedlacek@ericsson.com
Session Initiation Protocol (SIP) Response Code for Indication of Terminated Dialog
draft-ietf-sipcore-199-06.txt

Abstract

This specification defines a new Session Initiation Protocol (SIP) response code, 199 Early Dialog Terminated, that a SIP forking proxy and a User Agent Server (UAS) can use to indicate towards upstream SIP entities (including the User Agent Client (UAC)) that an early dialog has been terminated, before a final response is sent towards the SIP entities.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 4, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of
the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction .................................................. 3
2. Terminology .................................................. 4
3. Applicability and Limitation ................................. 4
4. User Agent Client behavior ................................... 5
5. User Agent Server behavior ................................... 6
6. Proxy behavior ................................................ 7
7. Backward compatibility ....................................... 9
8. Usage with SDP offer/answer ................................. 10
9. Message Flow Examples ....................................... 10
  9.1. Example with a forking proxy which generates 199 .... 10
  9.2. Example with a forking proxy which receives 200 OK ... 11
  9.3. Example with two forking proxies, of which one generates 199 ........................................... 12
10. Security Considerations ...................................... 13
11. IANA Considerations .......................................... 13
  11.1. IANA Registration of the 199 response code ........ 14
  11.2. IANA Registration of the 199 option-tag ............ 14
12. Acknowledgements ............................................ 14
13. Change Log .................................................... 14
14. References .................................................... 15
  14.1. Normative References .................................... 15
  14.2. Informational References ............................... 16
Author’s Address ................................................ 16
1. Introduction

As defined in RFC 3261 [RFC3261], a Session Initiation Protocol (SIP) early dialog is created when a non-100 provisional response is sent to the initial dialog initiation request (e.g. INVITE, outside an existing dialog). The dialog is considered to be in early state until a final response is sent.

When a proxy receives an initial dialog initiation request, it can forward the request towards multiple remote destinations. When the proxy does that, it performs forking [RFC3261].

When a forking proxy receives a non-100 provisional response, or a 2xx final response, it forwards the response upstream towards the sender of the associated request. After a forking proxy has forwarded a 2xx final response, it normally generates and sends CANCEL requests downstream towards all remote destinations where it previously forked the request associated with the 2xx final response and from which it has yet not received a final response. The CANCEL requests are sent in order to terminate any outstanding early dialogs associated with the request.

Upstream SIP entities might receive multiple 2xx final responses. When a SIP entity receives the first 2xx final response, and it does not intend to accept any subsequent 2xx final response, it will automatically terminate any other outstanding early dialog associated with the request. If the SIP entity receives a subsequent 2xx final response, it will normally generate and send an ACK request, followed with a BYE request, using the dialog identifier retrieved from the 2xx final response.

NOTE: A User Agent Client (UAC) can use the Request-Disposition header field [RFC3841] to request that proxies do not generate and send CANCEL requests downstream once they have received the first 2xx final response.

When a forking proxy receives a non-2xx final response, it does not always immediately forward the response upstream towards the sender of the associated request. Instead, the proxy "stores" the response and waits for subsequent final responses from other remote destinations where the associated request was forked. At some point the proxy uses a specified mechanism to determine the "best" final response code, and forwards a final response using that response code upstream towards the sender of the associated request. When an upstream SIP entity receives the non-2xx final response it will release resources associated with the session. The UAC will terminate, or retry, the session setup.
Since the forking proxy does not always immediately forward non-2xx final responses, upstream SIP entities (including the UAC that initiated the request) are not immediately informed that an early dialog has been terminated, and will therefore maintain resources associated with the early dialog reserved until a final response is sent by the proxy, even if the early dialog has already been terminated. A SIP entity could use the resources for other things, e.g. to accept subsequent early dialogs that it otherwise would reject.

This specification defines a new SIP response code, 199 Early Dialog Terminated. A forking proxy can send a 199 provisional response to inform upstream SIP entities that an early dialog has been terminated. A UAS can send a 199 response code, prior to sending a non-2xx final response, for the same purpose. SIP entities that receive the 199 response can use it to trigger the release of resources associated with the terminated early dialog. In addition, SIP entities might also use the 199 response to make policy related decisions related to early dialogs. For example, a media gate controlling SIP entity might use the 199 response when deciding for which early dialogs media will be passed.

Section 9 contains signalling examples that show when and how a forking proxy generates 199 responses in different situations.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

3. Applicability and Limitation

The 199 response code is an optimization, and it only optimizes how quickly recipients might be informed about terminated early dialogs. The achieved optimization is limited. Since the response is normally not sent reliably by an UAS, and can not be sent reliably when generated and sent by a proxy, it is possible that some or all of the 199 responses get lost before they reach the recipients. In such cases, recipients will behave the same as if the 199 response code were not used at all.

One example for which a UAC could use the 199 response, is that when it receives a 199 response it releases resources associated with the terminated early dialog. The UAC could also use the 199 response to make policy related decisions related to early dialogs. For example,
if a UAC is playing media associated with an early dialog, and it then receives a 199 response indicating the early dialog has been terminated, it could start playing media associated with a different early dialog.

Applications designers utilizing the 199 response code MUST ensure that the application’s user experience is acceptable if all 199 responses are lost, and not delivered to the recipients.

4. User Agent Client behavior

When a UAC sends an initial dialog initiation request, and if it is willing to receive 199 responses, it MUST insert a "199" option-tag in the Supported header field [RFC3261] of the request. The option-tag indicates that the UAC supports, and is willing to receive, 199 responses. A UAC SHOULD NOT insert a "199" option-tag in the Require or the Proxy-Require header field [RFC3261] of the request, since in many cases it would result in unnecessary session establishment failures.

NOTE: The UAC always needs to insert a "199" option-tag in the Supported header field, in order to indicate that it supports, and is willing to receive, 199 responses, even if it also inserts the option-tag in the Require or Proxy-Require header field.

It is RECOMMENDED that a UAC does not insert a "100rel" option-tag [RFC3262] in the Require header field when it also indicates support of 199 responses, unless the UAC also uses some other SIP extension or procedure that mandates it to do so. The reason is that proxies are not allowed to generate and send 199 responses when the UAC has required provisional responses to be sent reliably.

When a UAC receives a 199 response, it might release resources associated with the terminated early dialog. A UAC might also use the 199 response to make policy related decisions related to early dialogs.

NOTE: The 199 response indicates that the early dialog has been terminated, so there is no need for the UAC to send a BYE request in order to terminate the early dialog when it receives the 199 response.

NOTE: The 199 response does not affect other early dialogs associated with the session establishment. For those the normal SIP rules, regarding transaction timeout etc, still apply.

Once a UAC has received and accepted a 199 response, it MUST NOT send
Any media associated with the early dialog. In addition, if the UAC is able to associate received media with early dialogs, it MUST NOT process any received media associated with the early dialog that was terminated.

If multiple usages [RFC5057] are used within an early dialog, and it is not clear which dialog usage the 199 response terminates, SIP entities that keep dialog state SHALL NOT release resources associated with the early dialog when they receive the 199 response.

If a UAC receives an unreliably sent 199 response on a dialog which has not previously been established (this can happen if a 199 response reaches the client before the 18x response that would establish the early dialog) it SHALL discard the 199 responses. If a UAC receives a reliably sent 199 response on a dialog which has not previously been created, it MUST acknowledge the 199 response, as described in RFC 3262 [RFC3262].

If a UAC has received a 199 response for all early dialogs, and no early dialog associated session establishment remains, it maintains the "Proceeding" state [RFC3261] and waits for possible subsequent early dialogs to be established, and eventually for a final response to be received.

5. User Agent Server behavior

If a UAS receives an initial dialog initiation request, with a Supported header field that contains a "199" option-tag, it SHOULD NOT send a 199 response on an early dialog associated with the request, before it sends a non-2xx final response. Cases where a UAS might send a 199 response are if it has been configured to do so due to lack of support of the 199 response code by forking proxies or other intermediate SIP entities, or it is used in an environment that specifies that it shall send a 199 response before sending a non-2xx response.

NOTE: If a UAS has created multiple early dialogs associated with an initial dialog initiation request (the UAS is acting similar to a forking proxy), it does not always intend to send a final response on all of those early dialogs.

NOTE: If the Require header field of an initial dialog initiation request contains a "100rel" option-tag, proxies will not be able to generate and send 199 responses. In such cases the UAS might choose to send a 199 response on an early dialog, before it sends a non-2xx final response, even if it would not do so in other cases.
If the Supported header field of an initial dialog initiation request does not contain a "199" option-tag, the UAC MUST NOT send a 199 response on any early dialog associated with the request.

When a UAS generates a 199 response, the response MUST contain a To header field tag parameter [RFC3261], in order for other entities to identify the early dialog that has been terminated. The UAS MUST also insert a Reason header field [RFC3326] that contains a response code which describes the reason why the early dialog was terminated. The UAS MUST NOT insert a "199" option-tag in the Supported, Require or Proxy-Require header field of the 199 response.

If a UAS intends to send 199 responses, and if it supports the procedures defined in RFC 3840 [RFC3840], it MAY during the registration procedure use the sip.extensions feature tag [RFC3840] to indicate support of the 199 response code.

A 199 response SHOULD NOT contain an SDP offer/answer message body, unless required by the rules in RFC 3264 [RFC3264].

According to RFC 3264, if an INVITE request does not contain an SDP offer, and the 199 response is the first reliably sent response associated with the request, the 199 response is required to contain an SDP offer. In this case the UAS SHOULD send the 199 response unreliably, or send the 199 response reliably and include an SDP offer with no m- lines in the response.

Since a 199 response is only used for information purpose, the UAS SHOULD send it unreliably, unless the "100rel" option-tag is present in the Require header field of the associated request.

6. Proxy behavior

When a proxy receives a 199 response to an initial dialog initiation request, it MUST process the response as any other non-100 provisional response. The proxy will forward the response upstream towards the sender of the associated request. The proxy MAY release resources it has reserved associated with the early dialog that is terminated. If a proxy receives a 199 response out of dialog, it MUST process it as other non-100 provisional responses received out of dialog.

When a forking proxy receives a non-2xx final response to an initial dialog initiation request, that it recognizes as terminating one or more early dialogs associated with the request, it MUST generate and send a 199 response upstream for each of the terminated early dialogs that satisfy each of the following conditions:
- the forking proxy does not intend to forward the final response immediately (in accordance with rules for a forking proxy)

- the UAC has indicated support (by inserting the "199" option-tag in a Supported header field) of the 199 response code in the associated request

- the UAC has not required provisional responses to be sent reliably (by inserting the "100rel" option-tag in a Require or Proxy-Require header field) in the associated request

- the forking proxy has not already received and forwarded a 199 response for the early dialog

- the forking proxy has not already sent a final response for any of the early dialogs

As a consequence, once a final response to an initial dialog initiation request has been issued by the proxy, no further 199 responses associated with the request will be generated or forwarded by the proxy.

When a forking proxy forks an initial dialog initiation request, it generates a unique Via header branch parameter value for each forked leg. A proxy can determine whether additional forking has occurred downstream of the proxy by storing the top Via branch value from each response which creates an early dialog. If the same top Via branch value is received for multiple early dialogs, the proxy knows that additional forking has occurred downstream of the proxy. A non-2xx final response received for a specific early dialog also terminates all other early dialog for which the same top Via branch value was received in the responses which created those early dialogs.

Based on implementation policy, a forking proxy MAY wait before sending the 199 response, e.g. if it expects to receive a 2xx final response on another dialog shortly after it received the non-2xx final response which triggered the 199 response.

When a forking proxy generates a 199 response, the response MUST contain a To header field tag parameter, that identifies the terminated early dialog. A proxy MUST also insert a Reason header field that contains the SIP response code of the response that triggered the 199 response. The SIP response code in the Reason header field informs the receiver of the 199 response about the SIP response code that was used by the UAS to terminate the early dialog, and the receiver might use that information for triggering different types of actions and procedures. The proxy MUST NOT insert a "199" option-tag in the Supported, Require or Proxy-Require header field of
the 199 response.

A forking proxy that supports generating of 199 responses MUST keep track of early dialogs, in order to determine whether to generate a 199 response when the proxy receives a non-2xx final response. In addition, a proxy MUST keep track on which early dialogs it has received and forwarded 199 responses, in order to not generate additional 199 responses for those early dialogs.

If a forking proxy receives a reliably sent 199 response for a dialog, for which it has previously generated and sent a 199 response, it MUST forward the 199 response. If a proxy receives an unreliably sent 199 response, for which it has previously generated and sent a 199 response, it MAY forward the response, or it MAY discard it.

When a forking proxy generates and sends a 199 response, the response SHOULD NOT contain a Contact header field or a Record-Route header field [RFC3261].

If the Require header field of an initial dialog initiation request contains a "100rel" option-tag, a proxy MUST NOT generate and send 199 responses associated with that request. The reason is that a proxy is not allowed to generate and send 199 responses reliably.

7. Backward compatibility

Since all SIP entities involved in a session setup do not necessarily support the specific meaning of the 199 Early Dialog Terminated provisional response, the sender of the response MUST be prepared to receive SIP requests and responses associated with the dialog for which the 199 response was sent (a proxy can receive SIP messages from either direction). If such request is received by a UA, it MUST act in the same way as if it had received the request after sending the final non-2xx response to the INVITE request, as specified in RFC 3261. A UAC that receives a 199 response for an early dialog MUST NOT send any further requests on that dialog, except for requests which acknowledge reliable responses. A proxy MUST forward requests according to RFC 3261, even if the proxy has knowledge that the early dialog has been terminated.

A 199 response does not "replace" a final response. RFC 3261 specifies when a final response is sent.
8. Usage with SDP offer/answer

A 199 response SHOULD NOT contain an SDP offer/answer [RFC3264] message body, unless required by the rules in RFC 3264.

If an INVITE request does not contain an SDP offer, and the 199 response is the first reliably sent response, the 199 response is required to contain an SDP offer. In this case the UAS SHOULD send the 199 response unreliable, or include an SDP offer with no m- lines in a reliable 199 response.

9. Message Flow Examples

9.1. Example with a forking proxy which generates 199

The figure shows an example, where a proxy (P1) forks an INVITE received from UAC. The forked INVITE reaches UAS_2, UAS_3 and UAS_4, which send 18x provisional responses in order to establish early dialogs between themselves and the UAC. UAS_2 and UAS_3 reject the INVITE by sending a 4xx error response each. When P1 receives the 4xx responses it immediately sends 199 responses towards the UAC, to indicate that the early dialogs for which it received the 4xx responses have been terminated. The early dialog leg is shown in parenthesis.
The figure shows an example, where a proxy (P1) forks an INVITE request received from UAC. The forked request reaches UAS_2, UAS_3 and UAS_4, that all send 18x provisional responses in order to establish early dialogs between themselves and the UAC. Later UAS_4 accepts the session and sends a 200 OK final response. When P1 receives the 200 OK responses it immediately forwards it towards the UAC. P1 does not send 199 responses for the early dialogs from UAS_2 and UAS_3, since P1 has yet not received any final responses on those early dialogs (even if P1 sends CANCEL requests to UAS_2 and UAS_3 P1 may still receive 200 OK final response from UAS_2 or UAS_3, that P1 would have to forward towards the UAC. The early dialog leg is shown in parenthesis.
9.3. Example with two forking proxies, of which one generates 199

The figure shows an example, where a proxy (P1) forks an INVITE request received from UAC. One of the forked requests reaches UAS_2. The other requests reach another proxy (P2), that forks the request to UAS_3 and UAS_4. UAS_3 and UAS_4 send 18x provisional responses in order to establish early dialogs between themselves and UAC. Later UAS_3 and UAS_4 reject the INVITE request by sending a 4xx error response each. P2 does not support the 199 response code, and forwards a single 4xx response. P1 supports the 199 response code, and when it receives the 4xx response from P2, it also manages to associate the early dialogs from both UAS_3 and UAS_4 with the response. Therefore it generates and sends two 199 responses to indicate that the early dialogs from UAS_3 and UAS_4 have been terminated. The early dialog leg is shown in parenthesis.
10. Security Considerations

General security issues related to SIP responses are described in RFC 3261. Due to the nature of the 199 response, it may be attractive to use it for launching attacks in order to terminate specific early dialogs (other early dialogs will not be affected). In addition, if a man-in-the-middle generates and sends a 199 response, which terminates a specific dialog, towards the UAC, it can take a while until the UAS finds out that the UAC, and possible stateful intermediates, have terminated the dialog. SIP security mechanisms (e.g. hop-to-hop TLS) can be used to minimize, or eliminate, the risk for such attacks.

11. IANA Considerations

This section registers a new SIP response code and a new option-tag,
according to the procedures of RFC 3261.

11.1. IANA Registration of the 199 response code

This section registers a new SIP response code, 199. The required information for this registration, as specified in RFC 3261, is:

RFC Number: RFC XXXX [[NOTE TO IANA: Please replace XXXX with the RFC number of this specification]]

Response Code Number: 199

Default Reason Phrase: Early Dialog Terminated

11.2. IANA Registration of the 199 option-tag

This section registers a new SIP option-tag, 199. The required information for this registration, as specified in RFC 3261, is:

Name: 199

Description: This option-tag is for indicating support of the 199 Early Dialog Terminated provisional response code. When present in a Supported header of a request, it indicates that the UAC supports the 199 response code. When present in a Require or Proxy-Require header field of a request, it indicates that the UAS, or proxies, MUST support the 199 response code. It does not require the UAS, or proxies, to actually send 199 responses.

12. Acknowledgements

Thanks to Paul Kyzivat, Dale Worley, Gilad Shaham, Francois Audet, Attila Sipos, Robert Sparks, Brett Tate, Ian Elz, Hadriel Kaplan, Timothy Dwight, Dean Willis, Serhad Doken, John Elwell, Gonzalo Camarillo, Adam Roach, Bob Penfield, Tom Taylor, Ya Ching Tan, Keith Drage, Hans Erik van Elburg and Cullen Jennings for their feedback and suggestions.

13. Change Log

[RFC EDITOR NOTE: Please remove this section when publishing]

Changes from draft-ietf-sipcore-199-04
"Usage with 100rel" section removed based on comments from John Elwell (31.01.2011)

Editorial corrections based on comments from Paul Kyzivat (31.01.2011)

Changes from draft-ietf-sipcore-199-03
- RFC 3262 update removed
- Functional modification: proxy must not send 199 in case of Require:100rel
- Recommendation that UAC does not require reliable provisional responses with 199
- Clarification that Require:199 does not mandate the UAS to send a 199 response
- Clarification that a UAC needs to insert the 199 option-tag in a Supported header field, even if it also inserts the option-tag in a Require or Proxy-Require header field
- Editorial corrections

Changes from draft-ietf-sipcore-199-02
- Usage example section rewritten and clarified
- Requirement has been removed
- SIP has been added to document title
- Acronyms expanded in the abstract and throughout the document
- Editorial fixes throughout the document
- Indication added that document is aimed for standards track
- Some references made informative
- Additional text added regarding the usage of the Reason header
- SBC latching text has been removed
- Usage of Require/Proxy-Require header removed
- Additional text added regarding sending SDP offer in 199
- Note added, which clarifies that 199 does not affect other early dialogs
- References added to Security Considerations
- Clarification of local ringing tone
- Clarification that media must not be sent or processed after 199
- Text regarding sending media on terminated dialogs added to security section
- Change: UAS must send 199 reliably in case of Require:100rel
- Change: Section 4 of RFC 3262 updated

14. References

14.1. Normative References

14.2. Informational References


Author’s Address

Christer Holmberg
Ericsson
Hirsalantie 11
Jorvas 02420
Finland

Email: christer.holmberg@ericsson.com
Session Initiation Protocol (SIP) Event Notification Extension for 
Notification Rate Control 
draft-ietf-sipcore-event-rate-control-09

Abstract

This document specifies mechanisms for adjusting the rate of Session 
Initiation Protocol (SIP) event notifications. These mechanisms can 
be applied in subscriptions to all SIP event packages. This document 
updates RFC 3265.

Status of this Memo

This Internet-Draft is submitted in full conformance with the 
provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering 
Task Force (IETF). Note that other groups may also distribute 
working documents as Internet-Drafts. The list of current Internet-
Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months 
and may be updated, replaced, or obsoleted by other documents at any 
time. It is inappropriate to use Internet-Drafts as reference 
material or to cite them other than as "work in progress."

This Internet-Draft will expire on March 3, 2012.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the 
document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal 
Provisions Relating to IETF Documents 
(http://trustee.ietf.org/license-info) in effect on the date of 
publishation of this document. Please review these documents 
carefully, as they describe your rights and restrictions with respect 
to this document. Code Components extracted from this document must 
include Simplified BSD License text as described in Section 4.e of
the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.
1. Introduction

The SIP events framework [RFC3265] defines a generic framework for subscriptions to and notifications of events related to SIP systems. This framework defines the methods SUBSCRIBE and NOTIFY, and introduces the concept of an event package, which is a concrete application of the SIP events framework to a particular class of events.

One of the things the SIP events framework mandates is that each event package specification defines an absolute maximum on the rate at which notifications are allowed to be generated by a single notifier. Such a limit is provided in order to reduce network load.

All of the existing event package specifications include a maximum notification rate recommendation, ranging from once in every five seconds [RFC3856], [RFC3680], [RFC3857] to once per second [RFC3842]. Per the SIP events framework, each event package specification is also allowed to define additional throttle mechanisms which allow the subscriber to further limit the rate of event notification. So far none of the event package specifications have defined such a mechanism.

The resource list extension [RFC4662] to the SIP events framework also deals with rate limiting of event notifications. The extension allows a subscriber to subscribe to a heterogeneous list of resources with a single SUBSCRIBE request, rather than having to install a subscription for each resource separately. The event list subscription also allows rate limiting, or throttling of notifications, by means of the Resource List Server (RLS) buffering notifications of resource state changes, and sending them in batches. However, the event list mechanism provides no means for the subscriber to set the interval for the throttling.

Some event packages are also interested in specifying an absolute or an adaptive minimum rate at which notifications need to be generated by a notifier. This helps the subscriber to effectively use different trigger criterias within a subscription to eliminate unnecessary notifications but at the same time make sure that the current event state is periodically received.

This document defines an extension to the SIP events framework defining the following three Event header field parameters that allow a subscriber to set a maximum, a minimum and an adaptive minimum rate of notifications generated by the notifier:
max-rate: specifies a maximum number of notifications per second.

min-rate: specifies a minimum number of notifications per second.

adaptive-minimum-rate: specifies an adaptive minimum number of
notifications per second.

These mechanisms are applicable to any event subscription, both
single event subscription and event list subscription. A notifier
compliant to this specification will adjust the rate at which it
generates notifications.

2. Definitions and Document Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
"SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this
document are to be interpreted as described in RFC 2119 [RFC2119] and
indicate requirement levels for compliant implementations.

Indented passages such as this one are used in this document to
provide additional information and clarifying text. They do not
contain normative protocol behavior.

3. Overview

3.1. Use Case for Limiting the Maximum Rate of Notifications

A presence client in a mobile device contains a list of 100 buddies
or presentities. In order to decrease the processing and network
load of watching 100 presentities, the presence client has employed a
Resource List Server (RLS) with the list of buddies, and therefore
only needs a single subscription to the RLS in order to receive
notifications of the presence state of the resource list.

In order to control the buffer policy of the RLS, the presence client
sets a maximum rate of notifications. The RLS will buffer
notifications that are generated faster than they are allowed to be
sent due to the maximum rate and batch all of the buffered state
changes together in a single notification. The maximum rate applies
to the overall resource list, which means that there is a hard cap
imposed by the maximum rate to the number of notifications per second
the presence client can expect to receive.

The presence client can also modify the maximum rate of notifications
during the lifetime of the subscription. For example, if the mobile
device detects inactivity from the user for a period of time, the
presence client can simply pause notifications by choosing a "max-rate" parameter that allows only a single notification for the remainder of the subscription lifetime. When the user becomes active again, the presence client can resume the stream of notifications by re-subscribing with a "max-rate" parameter set to the earlier used value. Application of the mechanism defined by RFC 5839 [RFC5839] can also eliminate the transmission of a (full-state) notification carrying the latest resource state to the presence client after a subscription refresh.

3.2. Use Case for Setting a Minimum Rate for Notifications

A location application is monitoring the movement of a target. In order to decrease the processing and network load, the location application has made a subscription to a Location Server with a set of location filters [I-D.ietf-geopriv-loc-filters] that specify trigger criteria, e.g. to send an update only when the target has moved at least n meters. However, the application is also interested in receiving the current state periodically even if the state of the target has not changed enough to satisfy any of the trigger criteria, e.g., has not moved at least n meters within the period.

The location application sets a minimum rate of notifications and include it in the subscription sent to the Location Server. The "min-rate" parameter indicates the minimum number of notifications per second the notifier needs to generate.

The location application can also modify the minimum rate of notifications during the lifetime of the subscription. For example, when the subscription to the movement of a target is made, the notifier may not have the location information available. Thus, the first notification might be empty, or certain values might be absent. An important use case is placing constraints on when complete state should be provided after creating the subscription. Once state is acquired and the second notification is sent, the subscriber updates or changes the "min-rate" parameter to a more sensible value. This update can be performed in the response to the notification that contains the complete state information.

3.3. Use Case for Specifying an Adaptive Minimum Rate of Notifications

The minimum rate mechanism introduces a static and instantaneous rate control without the functionality to increase or decrease the notification rate adaptively. However, there are some applications that would work better with an adaptive minimum rate control.

A location application is monitoring the movement of a target. In order to decrease the processing in the application, the location
application wants to make a subscription that dynamically decreases the minimum rate of notifications if the target has sent out several notifications recently. However, if there have been only few recent notifications by the target, the location application wants the minimum rate of notifications to increase.

The location application sets an adaptive minimum rate of notifications and include it in the subscription sent to the Location Server. The "adaptive-min-rate" parameter value is used by the notifier to dynamically calculate the actual maximum time between two notifications. In order to dynamically calculate the maximum time, the notifier takes into consideration the rate at which notifications have been sent recently. In the adaptive minimum rate mechanism the notifier can increase or decrease the notification rate compared to the minimum rate mechanism based on the recent number of notifications sent out in the last period.

The location application can also modify the "adaptive-min-rate" parameter during the lifetime of the subscription.

3.4. Requirements

REQ1: The subscriber must be able to set a maximum rate of notifications in a specific subscription.

REQ2: The subscriber must be able to set a minimum rate of notifications in a specific subscription.

REQ3: The subscriber must be able to set an adaptive minimum rate of notifications in a specific subscription, which adjusts the minimum rate of notifications based on a moving average.

REQ4: It must be possible to apply the maximum rate, the minimum rate and the adaptive minimum rate mechanisms all together, or in any combination, in a specific subscription.

REQ5: It must be possible to use any of the different rate control mechanisms in subscriptions to any events.

REQ6: It must be possible to use any of the different rate control mechanisms together with any other event filtering mechanisms.

REQ7: The notifier must be allowed to use a policy in which the maximum rate, minimum rate and adaptive minimum rate parameters are adjusted from the value given by the subscriber.
For example, due to congestion reasons, local policy at
the notifier could temporarily dictate a policy that in
effect further decreases the maximum rate of
notifications. In another example, the notifier can
increase the subscriber proposed maximum rate so that at
least one notification is generated during the remainder
of the subscription lifetime.

REQ8: The different rate control mechanisms must address corner
cases for setting the notification rates appropriately. At a
minimum, the mechanisms must address the situation when the
time between two notifications would exceed the subscription
duration and should provide procedures for avoiding this
situation.

REQ9: The different rate control mechanisms must be possible to be
invoked, modified, or removed in the course of an active
subscription.

REQ10: The different rate control mechanisms must allow for the
application of authentication and integrity protection
mechanisms to subscriptions invoking that mechanism.

4. Basic Operations

4.1. Subscriber Behavior

In general, the way in which a subscriber generates SUBSCRIBE
requests and processes NOTIFY requests is according to RFC 3265
[RFC3265].

A subscriber that wants to have a maximum, minimum or adaptive
minimum rate of event notifications in a specific event subscription
does so by including a "max-rate", "min-rate" or "adaptive-min-rate"
Event header field parameter(s) as part of the SUBSCRIBE request.

A subscriber that wants to update a previously agreed event rate
control parameter does so by including the updated "max-rate", "min-
rate" or "adaptive-min-rate" Event header field parameter(s) as part
of a subsequent SUBSCRIBE request or a 2xx response to the NOTIFY
request. If the subscriber did not include at least one of the "max-
rate", "min-rate" or "adaptive-min-rate" header field parameters in
the most recent SUBSCRIBE request in a given dialog, the subscriber
MUST NOT include an Event header field with any of those parameters
in a 2xx response to a NOTIFY request in that dialog.
4.2. Notifier Behavior

In general, the way in which a notifier processes SUBSCRIBE requests and generates NOTIFY requests is according to RFC 3265 [RFC3265].

A notifier that supports the different rate control mechanisms MUST adjust its rate of notification according to the rate control values agreed with the subscriber. If the notifier needs to lower the subscription expiration value or if a local policy or other implementation-determined constraint at the notifier can not satisfy the rate control request, then the notifier can adjust (i.e. increase or decrease) appropriately the subscriber requested rate control values. The notifier MUST reflect back the possibly adjusted rate control values in a "max-rate", "min-rate" or "adaptive-min-rate" Subscription-State header field parameter of the subsequent NOTIFY requests.

5. Operation of the Maximum Rate Mechanism

5.1. Subscriber Behavior

A subscriber that wishes to apply a maximum rate to notifications in a subscription MUST construct a SUBSCRIBE request that includes the "max-rate" Event header field parameter. This parameter specifies the requested maximum number of notifications per second. The value of this parameter is a positive real number given by a finite decimal representation.

Note that the witnessed notification rate may not conform to the "max-rate" value for a number of reasons. For example, network jitter and retransmissions may result in the subscriber receiving the notifications more frequent than the "max-rate" value recommends.

A subscriber that wishes to update the previously agreed maximum rate of notifications MUST include the updated "max-rate" Event header field parameter in a subsequent SUBSCRIBE request or a 2xx response to the NOTIFY request.

A subscriber that wishes to remove the maximum rate control from notifications MUST indicate so by not including a "max-rate" Event header field parameter in a subsequent SUBSCRIBE request or a 2xx response to the NOTIFY request.

There are two main consequences for the subscriber when applying the maximum rate mechanism: state transitions may be lost and event notifications may be delayed. If either of these side effects
constitute a problem to the application that utilizes the event notifications, developers are instructed not to use the mechanism.

5.2. Notifier Behavior

A notifier that supports the maximum rate mechanism MUST extract the value of the "max-rate" Event header parameter from a SUBSCRIBE request or a 2xx response to the NOTIFY request and use it as the suggested maximum number of notifications per second. This value can be adjusted by the notifier, as defined in Section 5.3.

A compliant notifier MUST reflect back the possibly adjusted maximum rate of notifications in a "max-rate" Subscription-State header field parameter of the subsequent NOTIFY requests. The indicated "max-rate" value is adopted by the notifier, and the notification rate is adjusted accordingly.

A notifier that does not understand this extension will not reflect the "max-rate" Subscription-State header field parameter in the NOTIFY requests; the absence of this parameter indicates to the subscriber that no rate control is supported by the notifier.

A compliant notifier MUST NOT generate a notification if the interval since the most recent notification is less than the reciprocal value of the "max-rate" parameter, except when generating the notification either upon receipt of a SUBSCRIBE request, when the subscription state is changing from "pending" to "active" state or upon termination of the subscription (the last notification).

When a local policy dictates a maximum rate for notifications, a notifier will not generate notifications more frequently than the local policy maximum rate, even if the subscriber is not asking for maximum rate control. The notifier MAY inform the subscriber about such local policy maximum rate using the "max-rate" Subscription-State header field parameter included in subsequent NOTIFY requests.

Retransmissions of NOTIFY requests are not affected by the maximum rate mechanism, i.e., the maximum rate mechanism only applies to the generation of new transactions. In other words, the maximum rate mechanism does not in any way break or modify the normal retransmission mechanism specified in RFC 3261 [RFC3261].

5.3. Selecting the Maximum Rate

Special care needs to be taken when selecting the maximum rate. For example, the maximum rate could potentially set a minimum time value between notifications that exceeds the subscription expiration value. Such a configuration would effectively quench the notifier, resulting
in exactly two notifications to be generated. If the subscriber requests a maximum rate that would result in no notification before the subscription expiration, the notifier MUST increase the maximum rate and set it to the reciprocal value of the subscription expiration time left. According to RFC 3265 [RFC3265] the notifier may also shorten the subscription expiry anytime during an active subscription. If the subscription expiry is shortened during an active subscription, the notifier MUST also increase the "max-rate" value and set it to reciprocal value of the reduced subscription expiration time.

In some cases it makes sense to pause the notification stream on an existing subscription dialog on a temporary basis without terminating the subscription, e.g. due to inactivity on the application user interface. Whenever a subscriber discovers the need to perform the notification pause operation, it SHOULD set the maximum rate to the reciprocal value of the remaining subscription expiration value. This results in receiving no further notifications until the subscription expires or the subscriber sends a SUBSCRIBE request resuming notifications.

The notifier MAY decide to increase or decrease the proposed "max-rate" value by the subscriber based on its local policy, static configuration or other implementation-determined constraints. In addition, different event packages MAY define additional constraints for the allowed maximum rate ranges. Such constraints are out of the scope of this specification.

5.4. The Maximum Rate Mechanism for Resource List Server

When applied to a list subscription [RFC4662], the maximum rate mechanism has some additional considerations. Specifically, the maximum rate applies to the aggregate notification stream resulting from the list subscription, rather than explicitly controlling the notification of each of the implied constituent events. Moreover, the RLS can use the maximum rate mechanism on its own to control the rate of the back-end subscriptions to avoid overflowing its buffer.

The notifier is responsible for sending out event notifications upon state changes of the subscribed resource. We can model the notifier as consisting of four components: the event state resource(s), the Resource List Server (RLS) (or any other notifier), a notification buffer, and finally the subscriber, or watcher of the event state, as shown in Figure 1.
In short, the RLS reads event state changes from the event state resource, either by creating a back end subscription, or by other means; it packages them into event notifications and submits them into the output buffer. The rate at which this output buffer drains is controlled by the subscriber via the maximum rate mechanism. When a set of notifications are batched together, the way in which overlapping resource state is handled depends on the type of the resource state:

In theory, there are many buffer policies that the notifier could implement. However, we only concentrate on two practical buffer policies in this specification, leaving additional ones for further study and out of the scope of this specification. These two buffer policies depend on the mode in which the notifier is operating.
Full-state: Last (most recent) full state notification of each resource is sent out, and all others in the buffer are discarded. This policy applies to those event packages that carry full-state notifications.

Partial-state: The state deltas of each buffered partial notification per resource are merged, and the resulting notification is sent out. This policy applies to those event packages that carry partial-state notifications.

5.5. Buffer Policy Description

5.5.1. Partial State Notifications

With partial notifications, the notifier needs to maintain a separate buffer for each subscriber since each subscriber may have a different value for the maximum rate of notifications. The notifier will always need to keep both a copy of the current full state of the resource F, as well as the last successfully communicated full state view F’ of the resource in a specific subscription. The construction of a partial notification then involves creating a difference of the two states, and generating a notification that contains that difference.

When the maximum rate mechanism is applied to the subscription, it is important that F’ is replaced with F only when the difference of F and F’ was already included in a partial state notification to the subscriber allowed by the maximum rate mechanism. Additionally, the notifier implementation SHOULD check to see that the size of an accumulated partial state notification is smaller than the full state, and if not, the notifier SHOULD send the full state notification instead.

5.5.2. Full State Notifications

With full state notifications, the notifier only needs to keep the full state of the resource, and when that changes, send the resulting notification over to the subscriber.

When the maximum rate mechanism is applied to the subscription, the notifier receives the state changes of the resource, and generates a notification. If there is a pending notification, the notifier simply replaces that notification with the new notification, discarding the older state.
5.6. Estimated Bandwidth Savings

It is difficult to estimate the total bandwidth savings accrued by using the maximum rate mechanism over a subscription, since such estimates will vary depending on the usage scenarios. However, it is easy to see that given a subscription where several full state notifications would have normally been sent in any given interval set by the "max-rate" parameter, only a single notification is sent during the same interval when using the maximum rate mechanism yielding bandwidth savings of several times the notification size.

With partial-state notifications, drawing estimates is further complicated by the fact that the states of consecutive updates may or may not overlap. However, even in the worst case scenario, where each partial update is to a different part of the full state, a rate controlled notification merging all of these n partial states together should at a maximum be the size of a full-state update. In this case, the bandwidth savings are approximately n times the size of the header fields of the NOTIFY request.

It is also true that there are several compression schemes available that have been designed to save bandwidth in SIP, e.g., SigComp [RFC3320] and TLS compression [RFC3943]. However, such compression schemes are complementary rather than competing mechanisms to the maximum rate mechanism. After all, they can both be applied simultaneously.

6. Operation of the Minimum Rate Mechanism

6.1. Subscriber Behavior

A subscriber that wishes to apply a minimum rate to notifications in a subscription MUST construct a SUBSCRIBE request that includes the "min-rate" Event header field parameter. This parameter specifies the requested minimum number of notifications per second. The value of this parameter is a positive real number given by a finite decimal representation.

A subscriber that wishes to update the previously agreed minimum rate of notifications MUST include the updated "min-rate" Event header field parameter in a subsequent SUBSCRIBE request or a 2xx response to the NOTIFY request.

A subscriber that wishes to remove the minimum rate control from notifications MUST indicate so by not including a "min-rate" Event header field parameter in a subsequent SUBSCRIBE request or a 2xx response to the NOTIFY request.
The main consequence for the subscriber when applying the minimum rate mechanism is that it can receive a notification even if nothing has changed in the current state of the notifier. However, RFC 5839 [RFC5839] defines a mechanism that allows suppressing a NOTIFY request or a NOTIFY request body if the state has not changed.

6.2. Notifier Behavior

A notifier that supports the minimum rate mechanism MUST extract the value of the "min-rate" Event header field parameter from a SUBSCRIBE request or a 2xx response to the NOTIFY request and use it as the suggested minimum number of notifications per second. This value can be adjusted by the notifier, as defined in Section 6.3.

A compliant notifier MUST reflect back the possibly adjusted minimum rate of notifications in a "min-rate" Subscription-State header field parameter of the subsequent NOTIFY requests. The indicated "min-rate" value is adopted by the notifier, and the notification rate is adjusted accordingly.

A notifier that does not understand this extension, will not reflect the "min-rate" Subscription-State header field parameter in the NOTIFY requests; the absence of this parameter indicates to the subscriber that no rate control is supported by the notifier.

A compliant notifier MUST generate notifications when state changes occur or when the time since the most recent notification exceeds the reciprocal value of the "min-rate" parameter. Depending on the event package and subscriber preferences indicated in the SUBSCRIBE request, the NOTIFY request sent as a result of a minimum rate mechanism MUST contain either the current full state or the partial state showing the difference between the current state and the last successfully communicated state. If the subscriber and the notifier support the procedures in RFC 5839 [RFC5839] the complete NOTIFY request or the NOTIFY request body can be suppressed if the state has not changed from the previous notification.

Retransmissions of NOTIFY requests are not affected by the minimum rate mechanism, i.e., the minimum rate mechanism only applies to the generation of new transactions. In other words, the minimum rate mechanism does not in any way break or modify the normal retransmission mechanism.

6.3. Selecting the Minimum Rate

The minimum rate mechanism can be used to generate a lot of notifications, creating additional processing load for the notifier. Some of the notifications may also be unnecessary possibly repeating
already known state information to the subscriber. It is difficult
to provide generic guidelines for the acceptable minimum rate value
ranges, however the subscriber SHOULD request for the lowest possible
minimum rate. Different event packages MAY define additional
constraints for the allowed minimum rate values. Such constraints
are out of the scope of this specification.

The notifier MAY decide to increase or decrease the proposed "min-
rate" value by the subscriber based on its local policy, static
configuration or other implementation-determined constraints.

7. Operation of the Adaptive Minimum Rate Mechanism

7.1. Subscriber Behavior

A subscriber that wishes to apply an adaptive minimum rate to
notifications in a subscription MUST construct a SUBSCRIBE request
that includes the "adaptive-min-rate" Event header field parameter.
This parameter specifies an adaptive minimum number of notifications
per second. The value of this parameter is a positive real number
given by a finite decimal representation.

A subscriber that wishes to update the previously agreed adaptive
minimum rate of notifications MUST include the updated "adaptive-min-
rate" Event header field parameter in a subsequent SUBSCRIBE request
or a 2xx response to the NOTIFY request.

A subscriber that wishes to remove the adaptive minimum rate control
from notifications MUST indicate so by not including a "adaptive-min-
rate" Event header field parameter in a subsequent SUBSCRIBE request
or a 2xx response to the NOTIFY request.

The main consequence for the subscriber when applying the adaptive
minimum rate mechanism is that it can receive a notification even if
nothing has changed in the current state of the notifier. However,
RFC 5839 [RFC5839] defines a mechanism that allows suppressing a
NOTIFY request or a NOTIFY request body if the state has not changed.

7.2. Notifier Behavior

A notifier that supports the adaptive minimum rate mechanism MUST
extract the value of the "adaptive-min-rate" Event header parameter
from a SUBSCRIBE request or a 2xx response to the NOTIFY request and
use it to calculate the actual maximum time between two notifications
as defined in Section 7.4.

The "adaptive-min-rate" value can be adjusted by the notifier, as
defined in Section 7.3.

A compliant notifier MUST reflect back the possibly adjusted adaptive minimum rate of notifications in an "adaptive-min-rate" Subscription-State header field parameter of the subsequent NOTIFY requests. The indicated "adaptive-min-rate" value is adopted by the notifier, and the notification rate is adjusted accordingly.

A notifier that does not understand this extension will not reflect the "adaptive-min-rate" Subscription-State header parameter in the NOTIFY requests; the absence of this parameter indicates to the subscriber that no rate control is supported by the notifier.

A compliant notifier MUST generate notifications when state changes occur or when the time since the most recent notification exceeds the value calculated using the formula defined in Section 7.4. Depending on the event package and subscriber preferences indicated in the SUBSCRIBE request, the NOTIFY request sent as a result of a minimum rate mechanism MUST contain either the current full state or the partial state showing the difference between the current state and the last successfully communicated state. If the subscriber and the notifier support the procedures in RFC 5839 [RFC5839] the complete NOTIFY request or the NOTIFY request body can be suppressed if the state has not changed from the previous notification.

The adaptive minimum rate mechanism is implemented as follows:

1) When a subscription is first created, the notifier creates a record ("count" parameter) that keeps track of the number of notifications that have been sent in the "period". The "count" parameter is initialized to contain a history of having sent a "period" * adaptive-min-rate" number of notifications for the "period".

2) The "timeout" value is calculated according to the equation given in Section 7.4.

3) If the timeout period passes without a NOTIFY request being sent in the subscription, then the current resource state is sent (subject to any filtering associated with the subscription).

4) Whenever a NOTIFY request is sent (regardless of whether due to a "timeout" event or a state change), the notifier updates the notification history record stored in the "count" parameter, recalculates the value of "timeout," and returns to step 3.

Retransmissions of NOTIFY requests are not affected by the timeout, i.e., the timeout only applies to the generation of new transactions.
In other words, the timeout does not in any way break or modify the normal retransmission mechanism specified in RFC 3261 [RFC3261].

7.3. Selecting the Adaptive Minimum Rate

The adaptive minimum rate mechanism can be used to generate a lot of notifications, creating additional processing load for the notifier. Some of the notifications may also be unnecessary possibly repeating already known state information to the subscriber. It is difficult to provide generic guidelines for the acceptable adaptive minimum rate value ranges, however the subscriber SHOULD request for the lowest possible adaptive minimum rate value. Different event packages MAY define additional constraints for the allowed adaptive minimum rate values. Such constraints are out of the scope of this specification.

The notifier MAY decide to increase or decrease the proposed "adaptive-min-rate" value based on its local policy, static configuration or other implementation-determined constraints.

7.4. Calculating the Timeout

The formula used to vary the absolute pacing in a way that will meet the adaptive minimum rate requested over the period is given in equation (1):

\[
\text{timeout} = \frac{\text{count}}{(\text{adaptive-min-rate}^2) \times \text{period}}
\]  

(1)

The output of the formula, "timeout", is the time to the next notification, expressed in seconds. The formula has three inputs:

- adaptive-min-rate: The value of the "adaptive-min-rate" parameter conveyed in the Subscription-State header field.
- period: The rolling average period, in seconds. The granularity of the values for the "period" parameter is set by local policy at the notifier, however the notifier MUST choose a value greater than the reciprocal value of the "adaptive-min-rate" parameter. It is also RECOMMENDED that the notifier chooses a "period" parameter several times larger than reciprocal value of the "adaptive-min-rate" parameter in order to maximize the effectiveness of the equation (1). It is an implementation decision whether the notifier uses the same value of the "period" parameter for all subscriptions or individual values for each subscription.
count: The number of notifications that have been sent during the last "period" of seconds not including any retransmissions of requests.

In case both the maximum rate and the adaptive minimum rate mechanisms are used in the same subscription the formula used to dynamically calculate the timeout is given in equation (2):

\[
\text{timeout} = \text{MAX}[(1/\text{max-rate}), \text{count}/((\text{adaptive-min-rate} ^ 2)*\text{period})] \quad (2)
\]

max-rate: The value of the "max-rate" parameter conveyed in the Subscription-State header field.

The formula in (2) makes sure that for all the possible values of the "max-rate" and "adaptive-min-rate" parameters, with "adaptive-min-rate" < "max-rate", the timeout never results in a lower value than the reciprocal value of the "max-rate" parameter.

In some situation it may be beneficial for the notifier to achieve an adaptive minimum rate in a different way than the algorithm detailed in this document allows. However, the notifier MUST comply with any "max-rate" or "min-rate" parameters that have been negotiated.

8. Usage of the Maximum Rate, Minimum Rate and Adaptive Minimum Rate Mechanisms in a combination

Applications can subscribe to an event package using all the rate control mechanisms individually, or in combination; in fact there is no technical incompatibility among them. However there are some combinations of the different rate control mechanisms that make little sense to be used together. This section lists all the combinations that are possible to insert in a subscription; the utility to use each combination in a subscription is also analyzed.

maximum rate and minimum rate: This combination allows to reduce the notification rate, but at the same time assures the reception of periodic notifications.

A subscriber SHOULD choose a "min-rate" value lower than the "max-rate" value, otherwise the notifier MUST adjust the subscriber provided "min-rate" value to a value equal to or lower than the "max-rate" value.
maximum rate and adaptive minimum rate: It works in a similar way as the combination above, but with the difference that the interval at which notifications are assured changes dynamically.

A subscriber SHOULD choose a "adaptive-min-rate" value lower than the "max-rate" value, otherwise the notifier MUST adjust the subscriber provided "adaptive-min-rate" value to a value equal to or lower than the "max-rate" value.

minimum rate and adaptive minimum rate: When using the adaptive minimum rate mechanism, frequent state changes in a short period can result in no notifications for a longer period following the short period. The addition of the minimum rate mechanism ensures the subscriber always receives notifications after a specified interval.

A subscriber SHOULD choose a "min-rate" value lower than the "adaptive-min-rate" value, otherwise the notifier MUST NOT consider the "min-rate" value.

maximum rate, minimum rate and adaptive minimum rate: This combination makes little sense to be used although not forbidden.

A subscriber SHOULD choose a "min-rate" and "adaptive-min-rate" values lower than the "max-rate" value, otherwise the notifier MUST adjust the subscriber provided "min-rate" and "adaptive-min-rate" values to a value equal to or lower than the "max-rate" value.

A subscriber SHOULD choose a "min-rate" value lower than the "adaptive-min-rate" value, otherwise the notifier MUST NOT consider the "min-rate" value.

9. Protocol Element Definitions

This section describes the protocol extensions required for the different rate control mechanisms.

9.1. "max-rate", "min-rate" and "adaptive-min-rate" Header Field Parameters

The "max-rate", "min-rate" and "adaptive-min-rate" parameters are added to the rule definitions of the Event header field and the Subscription-State header field in RFC 3265 [RFC3265] grammar. Usage of this parameter is described in Section 5, Section 6 and Section 7.
9.2. Grammar

This section describes the Augmented BNF [RFC5234] definitions for the new header field parameters. Note that we derive here from the ruleset present in RFC 3265 [RFC3265], adding additional alternatives to the alternative sets of "event-param" and "subexp-params" defined therein.

\[
\begin{align*}
\text{event-param} & = \text{max-rate-param} \\
& / \text{min-rate-param} \\
& / \text{amin-rate-param} \\
\text{subexp-params} & = \text{max-rate-param} \\
& / \text{min-rate-param} \\
& / \text{amin-rate-param} \\
\text{max-rate-param} & = "\text{max-rate}\" \text{ EQUAL} \\
& (1*2\text{DIGIT} ["." 1*10\text{DIGIT}]) \\
\text{min-rate-param} & = "\text{min-rate}\" \text{ EQUAL} \\
& (1*2\text{DIGIT} ["." 1*10\text{DIGIT}]) \\
\text{amin-rate-param} & = "\text{adaptive-min-rate}\" \text{ EQUAL} \\
& (1*2\text{DIGIT} ["." 1*10\text{DIGIT}])
\end{align*}
\]

9.3. Event Header Field Usage in Responses to the NOTIFY request

This table expands the table described in Section 7.2 of RFC 3265 [RFC3265] allowing the Event header field to appear in a 2xx response to a NOTIFY request. The use of the Event header field in responses other than 2xx to NOTIFY requests is undefined and out of scope of this specification.

<table>
<thead>
<tr>
<th>Header field</th>
<th>where proxy</th>
<th>ACK</th>
<th>BYE</th>
<th>CAN</th>
<th>INV</th>
<th>OPT</th>
<th>REG</th>
<th>PRA</th>
<th>SUB</th>
<th>NOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event</td>
<td>2xx</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>o</td>
</tr>
</tbody>
</table>

A subscriber that wishes to update the previously agreed value for maximum, minimum or adaptive minimum rate of notifications MUST include all desired values for the "max-rate", "min-rate" and "adaptive-min-rate" parameters in an Event header field of the 2xx response to a NOTIFY request. Any of the other header field parameters currently defined for the Event header field by other specifications do not have a meaning if the Event header field is included in the 2xx response to the NOTIFY request. These header field parameters MUST be ignored by the notifier, if present.

The event type listed in the Event header field of the 2xx response to the NOTIFY request MUST match the event type of the Event header field in the corresponding NOTIFY request.
10. IANA Considerations

This specification registers three new SIP header field parameters, defined by the following information which is to be added to the Header Field Parameters and Parameter Values sub-registry under http://www.iana.org/assignments/sip-parameters.

<table>
<thead>
<tr>
<th>Header Field</th>
<th>Parameter Name</th>
<th>Predefined Values</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event</td>
<td>max-rate</td>
<td>No</td>
<td>[RFCxxxx]</td>
</tr>
<tr>
<td>Subscription-State</td>
<td>max-rate</td>
<td>No</td>
<td>[RFCxxxx]</td>
</tr>
<tr>
<td>Event</td>
<td>min-rate</td>
<td>No</td>
<td>[RFCxxxx]</td>
</tr>
<tr>
<td>Subscription-State</td>
<td>min-rate</td>
<td>No</td>
<td>[RFCxxxx]</td>
</tr>
<tr>
<td>Event</td>
<td>adaptive-min-rate</td>
<td>No</td>
<td>[RFCxxxx]</td>
</tr>
<tr>
<td>Subscription-State</td>
<td>adaptive-min-rate</td>
<td>No</td>
<td>[RFCxxxx]</td>
</tr>
</tbody>
</table>

(Note to the RFC Editor: please replace "xxxx" with the RFC number of this specification, when assigned.)

This specification also updates the reference defining the Event header field in the Header Fields sub-registry under http://www.iana.org/assignments/sip-parameters.

<table>
<thead>
<tr>
<th>Header Name</th>
<th>compact</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event</td>
<td>o</td>
<td>[RFC3265][RFCxxxx]</td>
</tr>
</tbody>
</table>

(Note to the RFC Editor: please replace "xxxx" with the RFC number of this specification, when assigned.)

11. Security Considerations

Naturally, the security considerations listed in RFC 3265 [RFC3265], which the rate control mechanisms described in this document extends, apply in entirety. In particular, authentication and message integrity SHOULD be applied to subscriptions with this extension.

RFC 3265 [RFC3265] recommends the integrity protection of the Event header field of SUBSCRIBE requests. Implementations of this extension SHOULD also provide integrity protection for the Event header field included in the 2xx response to the NOTIFY request. Without integrity protection an eavesdropper could see and modify the Event header field; or it could also manipulate the transmission of a 200 (OK) response to the NOTIFY request in order to suppress or flood notifications without the subscriber seeing what caused the problem.
When the maximum rate mechanism involves partial state notifications, the security considerations listed in RFC 5263 [RFC5263] apply in entirety.

12. Acknowledgments

Thanks to Pekka Pessi, Dean Willis, Eric Burger, Alex Audu, Alexander Milinski, Jonathan Rosenberg, Cullen Jennings, Adam Roach, Hisham Khartabil, Dale Worley, Martin Thomson, Byron Campen, Alan Johnston, Michael Procter, Janet Gunn and Ari Keranen for support and/or review of this work.

Thanks to Brian Rosen for the idea of the minimum and adaptive minimum rate mechanisms, and Adam Roach for the work on the algorithm for the adaptive minimum rate mechanism and other feedback.

13. References

13.1. Normative References


13.2. Informative References

[I-D.ietf-geopriv-loc-filters]  

[RFC3320]  

[RFC3680]  

[RFC3842]  

[RFC3856]  

[RFC3857]  

[RFC3943]  

[RFC5839]  

Authors’ Addresses

Aki Niemi  
Nokia  
P.O. Box 407  
NOKIA GROUP, FIN 00045  
Finland  

Phone: +358 50 389 1644  
Email: aki.niemi@nokia.com
Krisztian Kiss  
Nokia  
200 South Mathilda Ave  
Sunnyvale, CA  94086  
US  
Phone: +1 650 391 5969  
Email: krisztian.kiss@nokia.com

Salvatore Loreto  
Ericsson  
Hirsalantie 11  
Jorvas  02420  
Finland  
Email: salvatore.loreto@ericsson.com
Abstract

This specification defines a new Session Initiation Protocol (SIP) Via header field parameter, "keep", which allows adjacent SIP entities to explicitly negotiate usage of the Network Address Translation (NAT) keep-alive mechanisms defined in SIP Outbound, in cases where SIP Outbound is not supported, cannot be applied, or where usage of keep-alives is not implicitly negotiated as part of the SIP Outbound negotiation.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on July 24, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of
the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction ......................................................................................... 3
   1.1. Use-case: Dialog from non-registered UAs ................................. 3
   1.2. Use-case: SIP Outbound not supported ........................................ 3
   1.3. Use-case: SIP dialog initiated Outbound flows ............................ 3
2. Conventions ......................................................................................... 4
3. Definitions ........................................................................................... 4
4. User Agent and Proxy behavior ............................................................. 4
   4.1. General ......................................................................................... 4
   4.2. Lifetime of keep-alives .................................................................. 5
      4.2.1. General .................................................................................. 5
      4.2.2. Keep-alives associated with registration ............................... 5
      4.2.3. Keep-alives associated with dialog ....................................... 6
   4.3. Behavior of a SIP entity willing to send keep-alives .................... 6
   4.4. Behavior of a SIP entity willing to receive keep-alives ............... 7
5. Keep-alive frequency ............................................................................ 8
6. Connection reuse .................................................................................. 9
7. Examples ................................................................................................ 9
   7.1. General ......................................................................................... 9
    7.2. Keep-alive negotiation associated with registration:
      UA-proxy ......................................................................................... 10
    7.4. Keep-alive negotiation associated with dialog: UA-UA ......... 13
8. Grammar ............................................................................................... 14
   8.1. General ......................................................................................... 15
   8.2. ABNF ............................................................................................. 15
9. IANA Considerations ........................................................................... 15
   9.1. keep ............................................................................................... 15
10. Security Considerations ...................................................................... 15
11. Acknowledgements ............................................................................ 17
12. Change Log ......................................................................................... 17
13. References .......................................................................................... 18
   13.1. Normative References ................................................................. 18
   13.2. Informative References ............................................................... 18
Author's Address ..................................................................................... 19
1. Introduction

Section 3.5 of SIP Outbound [RFC5626] defines two keep-alive mechanisms. Even though the keep-alive mechanisms are separated from the rest of the SIP Outbound mechanism, SIP Outbound does not define a mechanism to explicitly negotiate usage of the keep-alive mechanisms. In some cases usage of keep-aliases can be implicitly negotiated as part of the SIP Outbound negotiation.

However, there are SIP Outbound use-cases where usage of keep-aliases is not implicitly negotiated as part of the SIP Outbound negotiation. In addition, there are cases where SIP Outbound is not supported, or where it cannot be applied, but where there is still a need to be able to negotiate usage of keep-aliases. Last, SIP Outbound only allows keep-aliases to be negotiated between a UA and an edge proxy, and not between other SIP entities.

This specification defines a new Session Initiation Protocol (SIP) [RFC3261] Via header field parameter, "keep", which allows adjacent SIP entities to explicitly negotiate usage of the NAT keep-alive mechanisms defined in SIP Outbound. The "keep" parameter allows SIP entities to indicate willingness to send keep-alives, to indicate willingness to receive keep-alives, and for SIP entities willing to receive keep-alives to provide a recommended keep-alive frequency.

The following sections describe use-cases where a mechanism to explicitly negotiate usage of keep-aliases is needed.

1.1. Use-case: Dialog from non-registered UAs

In some cases a User Agent Client (UAC) does not register itself before it establishes a dialog, but in order to maintain NAT bindings open during the lifetime of the dialog it still needs to be able to negotiate sending of keep-aliases towards its adjacent downstream SIP entity. A typical example is an emergency call, where a registration is not always required in order to make the call.

1.2. Use-case: SIP Outbound not supported

In some cases some SIP entities that need to be able to negotiate the use of keep-aliases might not support SIP Outbound. However, they might still support the keep-alive mechanisms defined in SIP Outbound, and need to be able to negotiate usage of them.

1.3. Use-case: SIP dialog initiated Outbound flows

SIP Outbound allows the establishment of flows using the initial request for a dialog. As specified in RFC 5626 [RFC5626], usage of
keep-alives is not implicitly negotiated for such flows.

2. Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14, RFC 2119 [RFC2119].

3. Definitions

Edge proxy: As defined in RFC 5626, a SIP proxy that is located topologically between the registering User Agent (UA) and the Authoritative Proxy.

NOTE: In some deployments the edge proxy might physically be located in the same SIP entity as the Authoritative Proxy.

Keep-alives: The keep-alive messages defined in RFC 5626.

"keep" parameter: A SIP Via header field parameter that a SIP entity can insert in the topmost Via header field that it adds to the request, to explicitly indicate willingness to send keep-alives towards its adjacent downstream SIP entity. A SIP entity can add a parameter value to the "keep" parameter in a response to explicitly indicate willingness to receive keep-alives from its adjacent upstream SIP entity.

SIP entity: SIP User Agent (UA), or proxy, as defined in RFC 3261.

Adjacent downstream SIP entity: The adjacent SIP entity in the direction towards which a SIP request is sent.

Adjacent upstream SIP entity: The adjacent SIP entity in the direction from which a SIP request is received.

4. User Agent and Proxy behavior

4.1. General

This section describes how SIP UAs and proxies negotiate usage of keep-alives associated with a registration, or a dialog, which types of SIP requests can be used in order to negotiate the usage, and the lifetime of the negotiated keep-alives.
SIP entities indicate willingness to send keep-alives towards the adjacent downstream SIP entity using SIP requests. The associated responses are used by SIP entities to indicate willingness to receive keep-alives. SIP entities that indicate willingness to receive keep-alives can provide a recommended keep-alive frequency.

The procedures to negotiate usage of keep-alives are identical for SIP UAs and proxies.

In general, it can be useful for SIP entities to indicate willingness to send keep-alives, even if they are not aware of any necessity for them to send keep-alives, since the adjacent downstream SIP entity might have knowledge about the necessity. Similarly, if the adjacent upstream SIP entity has indicated willingness to send keep-alives, it can be useful for SIP entities to indicate willingness to receive keep-alives, even if they are not aware of any necessity for the adjacent upstream SIP entity to send them.

NOTE: Usage of keep-alives is negotiated per direction. If a SIP entity has indicated willingness to receive keep-alives from an adjacent SIP entity, sending of keep-alives towards that adjacent SIP entity needs to be separately negotiated.

NOTE: Since there are SIP entities that already use a combination of Carriage Return and Line Feed (CRLF) as keep-alive messages, and SIP entities are expected to be able to receive those, this specification does not forbid the sending of double-CRLF keep-alive messages towards an adjacent SIP entity even if usage of keep-alives with that SIP entity has not been negotiated. However, the "keep" parameter is still important in order for a SIP entity to indicate that it supports sending of double-CRLF keep-alive messages, so that the adjacent downstream SIP entity does not use other mechanisms (e.g. short registration refresh intervals) in order to keep NAT bindings open.

4.2. Lifetime of keep-alives

4.2.1. General

The lifetime of negotiated keep-alives depends on whether the keep-alives are associated with a registration or a dialog. This section describes the lifetime of negotiated keep-alives.

4.2.2. Keep-alives associated with registration

SIP entities use a registration request in order to negotiate usage of keep-alives associated with a registration. Usage of keep-alives can be negotiated when the registration is established, or later
during the registration. Once negotiated, keep-alives are sent until
the registration is terminated, or until a subsequent registration
refresh request is sent or forwarded. When a subsequent registration
refresh request is sent or forwarded, if a SIP entity is willing to
continue sending keep-alives associated with the registration, usage
of keep-alives MUST be re-negotiated. If usage is not successfully
re-negotiated, the SIP entity MUST cease sending of keep-alives
associated with the registration.

NOTE: Sending of keep-alives associated with a registration can only
be negotiated in the direction from the registering SIP entity
towards the registrar.

4.2.3. Keep-alives associated with dialog

SIP entities use an initial request for a dialog, or a mid-dialog
target refresh request [RFC3261], in order to negotiate sending and
receiving of keep-alives associated with a dialog. Usage of keep-
alives can be negotiated when the dialog is established, or later
during the lifetime of the dialog. Once negotiated, keep-alives MUST
be sent for the lifetime of the dialog, until the dialog is
terminated. Once usage of keep-alives associated with a dialog has
been negotiated, it is not possible to re-negotiate the usage
associated with the dialog.

4.3. Behavior of a SIP entity willing to send keep-alives

As defined in RFC 5626, a SIP entity that supports sending of keep-
alives must act as a Session Traversal Utilities for NAT (STUN)
client [RFC5389]. The SIP entity must support those aspects of STUN
that are required in order to apply the STUN keep-alive mechanism
defined in RFC 5626, and it must support the CRLF keep-alive
mechanism defined in RFC 5626. RFC 5626 defines when to use STUN,
respectively double-CRLF, for keep-alives.

When a SIP entity sends or forwards a request, if it wants to
negotiate the sending of keep-alives associated with a registration,
or a dialog, it MUST insert a "keep" parameter in the topmost Via
header field that it adds to the request, to indicate willingness to
send keep-alives.

When the SIP entity receives the associated response, if the "keep"
parameter in the topmost Via header field of the response contains a
"keep" parameter value, it MUST start sending keep-alives towards the
same destination where it would send a subsequent request (e.g.
REGISTER requests and initial requests for dialog) associated with
the registration (if the keep-alive negotiation is for a
registration), or where it would send subsequent mid-dialog requests
(if the keep-alive negotiation is for a dialog). Subsequent mid-dialog requests are addressed based on the dialog route set.

Once a SIP entity has negotiated sending of keep-alives associated with a dialog towards an adjacent SIP entity, it MUST NOT insert a "keep" parameter in any subsequent SIP requests, associated with the dialog, towards that adjacent SIP entity. Such "keep" parameter MUST be ignored, if received.

Since an ACK request does not have an associated response, it cannot be used to negotiate usage of keep-alives. Therefore, a SIP entity MUST NOT insert a "keep" parameter in the topmost Via header field of an ACK request. Such "keep" parameter MUST be ignored, if received.

A SIP entity MUST NOT indicate willingness to send keep-alives associated with a dialog, unless it has also inserted itself in the dialog route set [RFC3261].

NOTE: When a SIP entity sends an initial request for a dialog, if the adjacent downstream SIP entity does not insert itself in the dialog route set using a Record-Route header field [RFC3261], the adjacent downstream SIP entity will change once the dialog route set has been established. If a SIP entity inserts a "keep" parameter in the topmost Via header field of an initial request for a dialog, and the "keep" parameter in the associated response does not contain a parameter value, the SIP entity might choose to insert a "keep" parameter in the topmost Via header field of a subsequent SIP request associated with the dialog, in case the new adjacent downstream SIP entity (based on the dialog route set) is willing to receive keep-alives (in which case it will add a parameter value to the "keep" parameter).

If an INVITE request is used to indicate willingness to send keep-alives, as long as at least one response (provisional or final) to the INVITE request contains a "keep" parameter with a parameter value, it is seen as an indication that the adjacent downstream SIP entity is willing to receive keep-alives associated with the dialog on which the response is received.

4.4. Behavior of a SIP entity willing to receive keep-alives

As defined in RFC 5626, a SIP entity that supports receiving of keep-alives must act as a STUN server [RFC5389]. The SIP entity must support those aspects of STUN that are required in order to apply the STUN keep-alive mechanism defined in RFC 5626, and it must support the CRLF keep-alive mechanism defined in RFC 5626.

When a SIP entity sends or forwards a response, and the adjacent
upstream SIP entity indicated willingness to send keep-alives, if the SIP entity is willing to receive keep-alives associated with the registration, or the dialog, from the adjacent upstream SIP entity it MUST add a parameter value to the "keep" parameter, before sending or forwarding the response. The parameter value, if present and with a value other than zero, represents a recommended keep-alive frequency, given in seconds.

There might be multiple responses to an INVITE request. When a SIP entity indicates willingness to receive keep-alives in a response to an INVITE request, it MUST add a parameter value to the "keep" parameter in at least one reliable response to the request. The SIP entity MAY add identical parameter values to the "keep" parameters in other responses to the same request. The SIP entity MUST NOT add different parameter value to the "keep" parameters in responses to the same request. The SIP entity SHOULD indicate the willingness to receive keep-alives as soon as possible.

A SIP entity MUST NOT indicates willingness to receive keep-alives associated with a dialog, unless it has also inserted itself in the dialog route set [RFC3261].

5. Keep-alive frequency

If a SIP entity receives a SIP response, where the topmost Via header field contains a "keep" parameter with a non-zero value that indicates a recommended keep-alive frequency, given in seconds, it MUST use the procedures defined for the Flow-Timer header field [RFC5626]. According to the procedures, the SIP entity must send keep-alives at least as often as the indicated recommended keep-alive frequency, and if the SIP entity uses the recommended keep-alive frequency then it should send its keep-alives so that the interval between each keep-alive is randomly distributed between 80% and 100% of the recommended keep-alive frequency.

If the received "keep" parameter value is zero, the SIP entity can send keep-alives at its discretion. RFC 5626 provides additional guidance on selecting the keep-alive frequency in case a recommended keep-alive frequency is not provided.

This specification does not specify actions to take if negotiated keep-alives are not received. As defined in RFC 5626, the receiving SIP entity may consider a connection to be dead in such situations.

If a SIP entity that adds a parameter value to the "keep" parameter, in order to indicate willingness to receive keep-alives, also inserts a Flow-Timer header field (that can happen if the SIP entity is using
both the Outbound mechanism and the keep-alive mechanism) in the same SIP message, the header field value and the "keep" parameter value MUST be identical.

SIP Outbound uses the Flow-Timer header field to indicate the server-recommended keep-alive frequency. However, it will only be sent between a UA and an edge proxy. Using the "keep" parameter, however, the sending and receiving of keep-alives might be negotiated between multiple entities on the signalling path. In addition, since the server-recommended keep-alive frequency might vary between different SIP entities, a single Flow-Timer header field can not be used to indicate all the different frequency values.

6. Connection reuse

Keep-alives are often sent in order to keep NAT bindings open, so that the NAT may be passed by SIP requests sent in the reverse direction, reusing the same connection, or for non-connection-oriented transport protocols, reusing the same path. This specification does not define such connection reuse mechanism. The keep-alive mechanism defined in this specification is only used to negotiate the sending and receiving of keep-alives. Entities that want to reuse connections need to use another mechanism to ensure that security aspects associated with connection reuse are taken into consideration.

RFC 5923 [RFC5923] specifies a mechanism for using connection-oriented transports to send requests in the reverse direction, and an entity that wants to use connection-reuse as well as indicate support of keep-alives on that connection will insert both the "alias" parameter defined in RFC 5923 as well as the "keep" parameter defined in this specification.

SIP Outbound specifies how registration flows are used to send requests in the reverse direction.

7. Examples

7.1. General

This section shows example flows where usage of keep-alives, associated with a registration and a dialog, is negotiated between different SIP entities.

NOTE: The examples do not show the actual syntactical encoding of the request lines, response lines and the Via header fields, but rather a
pseudo code in order to identity the message type and to which SIP entity a Via header field is associated.

7.2. Keep-alive negotiation associated with registration: UA-proxy

Figure 1 shows an example where Alice sends an REGISTER request. She indicates willingness of sending keep-alive by inserting a "keep" parameter in her Via header field of the request. The edge proxy (P1) forwards the request towards the registrar.

P1 is willing to receive keep-alives from Alice for the duration of the registration, so when P1 receives the associated response it adds a "keep" parameter value, which indicates a recommended keep-alive frequency of 30 seconds, to Alice’s Via header field, before it forwards the response towards Alice.

When Alice receives the response, she determines from her Via header field that P1 is willing to receive keep-alives associated with the registration. Until the registration expires, or Alice sends a registration refresh request, Alice then sends periodic keep-alives (in this example using the STUN keep-alive technique) towards P1, using the recommended keep-alive frequency indicated by the "keep" parameter value.
### 7.3. Keep-alive negotiation associated with dialog: UA-proxy

Figure 2 shows an example where Alice sends an initial INVITE request for a dialog. She indicates willingness to send keep-alive by inserting a "keep" parameter in her Via header field of the request. The edge proxy (P1) adds itself to the dialog route set by adding itself to a Record-Route header field, before it forwards the request towards Bob.

P1 is willing to receive keep-alives from Alice for the duration of the dialog, so when P1 receives the associated response it adds a "keep" parameter value, which indicates a recommended keep-alive frequency of 30 seconds, to Alice’s Via header field, before it forwards the response towards Alice.

When Alice receives the response, she determines from her Via header field that P1 is willing to receive keep-alives associated with the dialog. For the lifetime of the dialog, Alice then sends periodic keep-alives to P1.

---

<table>
<thead>
<tr>
<th>Alice</th>
<th>P1</th>
<th>REGISTRAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>--- REGISTER--- &gt;</td>
<td>--- REGISTER--- &gt;</td>
<td>--- REGISTER--- &gt;</td>
</tr>
<tr>
<td>Via: Alice;keep</td>
<td>Via: P1</td>
<td>Via: Alice;keep</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;-- 200 OK ---&quot;</td>
<td>&quot;-- 200 OK ---&quot;</td>
<td>&quot;-- 200 OK ---&quot;</td>
</tr>
<tr>
<td>Via: P1</td>
<td>Via: P1</td>
<td>Via: Alice;keep</td>
</tr>
<tr>
<td>&quot;-- 200 OK ---&quot;</td>
<td>&quot;-- 200 OK ---&quot;</td>
<td>&quot;-- 200 OK ---&quot;</td>
</tr>
<tr>
<td>Via: Alice;keep=30</td>
<td>Via: Alice;keep=30</td>
<td>Via: Alice;keep=30</td>
</tr>
</tbody>
</table>

*** Timeout ***

*** Timeout ***

---

Figure 1: Example call flow
keep-alives (in this example using the STUN keep-alive technique) towards P1, using the recommended keep-alive frequency indicated by the "keep" parameter value.

Alice                        P1                         Bob

--- INVITE -------------->  --- INVITE -------------->  --- INVITE -------------->
Via: Alice;keep            Via: P1
Via: Alice;keep            Via: Alice;keep
Record-Route: P1           Record-Route: P1

<-- 200 OK ---------------
Via: P1
Via: Alice;keep
Record-Route: P1

--- ACK ------------------>  --- ACK ------------------>  --- ACK ------------------>

*** Timeout ***

=== STUN request ========
<== STUN response ========

*** Timeout ***

=== STUN request ========
<== STUN response ========

--- BYE ------------------>
--- BYE ------------------>
<-- 200 OK ---------------

Figure 2: Example call flow
7.4. Keep-alive negotiation associated with dialog: UA-UA

Figure 3 shows an example where Alice sends an initial INVITE request for a dialog. She indicates willingness to send keep-alive by inserting a "keep" parameter in her Via header field of the request. The edge proxy (P1) does not add itself to the dialog route set, by adding itself to a Record-Route header field, before it forwards the request towards Bob.

When Alice receives the response, she determines from her Via header field that P1 is not willing to receive keep-alives associated with the dialog from her. When the dialog route set has been established, Alice sends a mid-dialog UPDATE request towards Bob (since P1 did not insert itself in the dialog route set), and she once again indicates willingness to send keep-alives by inserting a "keep" parameter in her Via header field of the request. Bob supports the keep-alive mechanism, and is willing to receive keep-alives associated with the dialog from Alice, so he creates a response and adds a "keep" parameter value, which indicates a recommended keep-alive frequency of 30 seconds, to Alice’s Via header field, before he forwards the response towards Alice.

When Alice receives the response, she determines from her Via header field that Bob is willing to receive keep-alives associated with the dialog. For the lifetime of the dialog, Alice then sends periodic keep-alives (in this example using the STUN keep-alive technique) towards Bob, using the recommended keep-alive frequency indicated by the "keep" parameter value.
Figure 3: Example call flow

8. Grammar
8.1. General

This section describes the syntax extensions to the ABNF syntax defined in RFC 3261, by defining a new Via header field parameter, "keep". The ABNF defined in this specification is conformant to RFC 5234 [RFC5234].

8.2. ABNF

via-params =/ keep

keep = "keep" [ EQUAL 1*(DIGIT) ]

9. IANA Considerations

9.1. keep

This specification defines a new Via header field parameter called keep in the "Header Field Parameters and Parameter Values" sub-registry as per the registry created by [RFC3968]. The syntax is defined in Section 8. The required information is:

<table>
<thead>
<tr>
<th>Header Field</th>
<th>Parameter Name</th>
<th>Predefined Values</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Via</td>
<td>keep</td>
<td>No</td>
<td>[RFCXXXX]</td>
</tr>
</tbody>
</table>

10. Security Considerations

SIP entities that send or receive keep-alives are often required to use a connection reuse mechanism, in order to ensure that requests sent in the reverse direction, towards the sender of the keep-alives, traverse NATs etc. This specification does not specify a connection reuse mechanism, and it does not address security issues related to connection reuse. SIP entities that wish to reuse connections need to use a dedicated connection reuse mechanism, in conjunction with the keep-alive negotiation mechanism.

Unless SIP messages are integrity protected hop-by-hop, e.g. using Transport Layer Security (TLS) [RFC5246] or Datagram Transport Layer Security (DTLS) [RFC4347], a man-in-the-middle can modify Via header fields used by two entities to negotiate sending of keep-alives, e.g.
by removing the indications used to indicate willingness to send and receive keep-alives, or by decreasing the timer value to a very low value, which might trigger additional resource consumption due to the frequently sent keep-alives.

The behavior defined in Sections 4.3 and 4.4 require a SIP entity using the mechanism defined in this specification to place a value in the "keep" parameter in the topmost Via header field value of a response the SIP entity sends. They do not instruct the entity to place a value in a "keep" parameter of any request it forwards. In particular, SIP proxies MUST NOT place a value into the keep parameter of the topmost Via header field value of a request it receives before forwarding it. A SIP proxy implementing this specification SHOULD remove any keep parameter values in any Via header field values below the topmost one in responses it receives before forwarding them.

When requests are forwarded across multiple hops, it is possible for a malicious downstream SIP entity to tamper with the accrued values in the Via header field. The malicious SIP entity could place a value, or change an existing value in a "keep" parameter in any of the Via header field values, not just the topmost value. A proxy implementation that simply forwards responses by stripping the topmost Via header field value and not inspecting the resulting new topmost Via header field value risks being adversely affected by such a malicious downstream SIP entity. In particular, such a proxy may start receiving STUN requests if it blindly forwards a response with a keep parameter with a value it did not create in the topmost Via header field.

To lower the chances of the malicious SIP entity’s actions having adverse effects on such proxies, when a SIP entity sends STUN keep-alives to an adjacent downstream SIP entity and does not receive a response to those STUN messages, it MUST, based on the procedure in section 4.4.2 of RFC 5626, after 7 retransmissions, or when an error response is received for the STUN request, stop sending keep-alives for the remaining duration of the dialog (if the sending of keep-alives were negotiated for a dialog) or until the sending of keep-alives is re-negotiated for the registration (if the sending keep-alives were negotiated for a registration).

Apart from the issues described above, this specification does not introduce security considerations in addition to those specified for keep-alives in [RFC5626].
11. Acknowledgements

Thanks to Staffan Blau, Francois Audet, Hadriel Kaplan, Sean Schneyer and Milo Orsic for their comments on the initial draft. Thanks to Juha Heinaenen, Jiri Kuthan, Dean Willis, John Elwell, Paul Kyzivat, Peter Musgrave, Dale Worley, Adam Roach and Robert Sparks for their comments on the list. Thanks to Vijay Gurbani for providing text about the relationship with the connect reuse specification.

12. Change Log

[RFC EDITOR NOTE: Please remove this section when publishing]

Changes from draft-ietf-sipcore-keep-11
   - Editorial fixes based on last call comments by Peter Saint-Andre (Jan 11th)
   - TLS and DTLS references added
   - Clarification that the sending of keep-alives stops after 7 retransmissions
   - Editorial fixes based on last call comments by Alexey Melnikov (Jan 12th)
   - Additional text added to Grammar section
   - Editorial fixes based on last call comments by Adrian Farrel (Jan 16th)
   - Editorial fixes based on last call comments by Sean Turner (Jan 20th)
   - Reference clean-ups

Changes from draft-ietf-sipcore-keep-10
   - Editorial fixes based on last call comments by Juergen Schoenwaelder (Dec 21st)
   - Editorial fixes based on last call comments by Roni Even (Dec 28th)

Changes from draft-ietf-sipcore-keep-09
   - Changes based on AD review comments by Robert Sparks
   - Redundant paragraph removed from security considerations

Changes from draft-ietf-sipcore-keep-08
   - Changes based on AD review comments by Robert Sparks
   - Additional security considerations text provided by Robert Sparks

Changes from draft-ietf-sipcore-keep-07
13. References

13.1. Normative References


13.2. Informative References


Author’s Address

Christer Holmberg
Ericsson
Hirsalantie 11
Jorvas 02420
Finland

Email: christer.holmberg@ericsson.com
Abstract

This document defines an extension to the Session Initiation Protocol (SIP) to convey geographic location information from one SIP entity to another SIP entity. The extension covers end-to-end conveyance as well as location-based routing, where SIP intermediaries make routing decisions based upon the location of the user agent client.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on April 25, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document.
Table of Contents

1. Conventions and Terminology used in this document ........ 3
2. Introduction ............................................. 3
3. Overview of SIP Location Conveyance .......................... 4
   3.1 Location Conveyed by Value ............................ 4
   3.2 Location Conveyed as a Location URI ...................... 4
   3.3 Location Conveyed though a SIP Intermediary .............. 5
   3.4 SIP Intermediary Replacing Bad Location ................... 6
4. SIP Modifications for Geolocation Conveyance .................. 8
   4.1 The Geolocation Header .................................. 8
   4.2 424 (Bad Location Information) Response Code ............ 10
   4.3 The Geolocation-Error Header ............................. 11
   4.4 The 'geolocation' Option Tag ............................. 14
   4.5 Location URIs in Message Bodies ........................ 14
   4.6 Location URIs Allowed .................................. 14
5. Geolocation Examples ....................................... 14
   5.1 Location-by-value (Coordinate Format) ..................... 14
   5.2 Two Locations Composed in Same Location Object Example ... 16
6. Geopriv Privacy Considerations ............................... 18
7. Security Considerations ...................................... 18
8. IANA Considerations ......................................... 19
   8.1 IANA Registration for New SIP Geolocation Header ........ 20
   8.2 IANA Registration for New SIP 'geolocation' Option Tag ... 20
   8.3 IANA Registration for New 424 Response Code ............ 20
   8.4 IANA Registration for New SIP Geolocation-Error Header ... 20
   8.5 IANA Registration for New SIP Geolocation-Error Codes ... 20
   8.6 IANA Registration of Location URI Schemes ................ 21
9. Acknowledgements ........................................... 21
10. References .................................................. 22
   10.1 Normative References .................................... 22
   10.2 Informative References ................................. 23
1. Conventions and Terminology used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119]. This document furthermore uses numerous terms defined in RFC 3693 [RFC3693], including Location Object, Location Recipient, Location Server, Target, and Using Protocol.

2. Introduction

Session Initiation Protocol (SIP) [RFC3261] creates, modifies and terminates multimedia sessions. SIP carries certain information related to a session while establishing or maintaining calls. This document defines how SIP conveys geographic location information of a Target (Target) to a Location Recipient (LR). SIP acts as a Using Protocol of location information, as defined in RFC 3693.

In order to convey location information, this document specifies a new SIP header, the Geolocation header, which carries a reference to a Location Object. That Location Object may appear in a MIME body attached to the SIP request, or it may be a remote resource in the network.

Note that per RFC 3693, a Target is an entity whose location is being conveyed. Thus, a Target could be a SIP user agent (UA), some other IP device (a router or a PC) that does not have a SIP stack, a non-IP device (a person or a black phone) or even a non-communications device (a building or store front). In no way does this document assume that the SIP user agent client which sends a request containing a location object is necessarily the Target. The location of a Target conveyed within SIP typically corresponds to that of a device controlled by the Target, for example, a mobile phone, but such devices can be separated from their owners, and moreover, in some cases the user agent may not know its own location.

In the SIP context, a location recipient will most likely be a SIP UA, but due to the mediated nature of SIP architectures, location information conveyed by a single SIP request may have multiple recipients, as any SIP proxy server in the signaling path that inspects the location of the Target must also be considered a Location Recipient. In presence-like architectures, an intermediary that receives publications of location information and distributes them to watchers acts as a Location Server per RFC 3693. This location conveyance mechanism can also be used to deliver URIs point to such Location Servers where prospective Location Recipients can request Location Objects.
3. Overview of SIP Location Conveyance

An operational overview of SIP location conveyance can be shown in 4 basic diagrams, with most applications falling under one of the following basic use cases. Each is separated into its own subsection here in section 3.

Each diagram has Alice and Bob as UAs. Alice is the Target, and Bob is an LR. A SIP intermediary appears in some of the diagrams. Any SIP entity that receives and inspects location information is an LR, therefore any of the diagrams the SIP intermediary receives the SIP request is potentially an LR - though that does not mean such an intermediary necessarily has to route the SIP request based on the location information. In some use cases, location information passes through the LS on the right of each diagram.

3.1 Location Conveyed by Value

We start with the simplest diagram of Location Conveyance, Alice to Bob, where no other layer 7 entities are involved.

```
Alice | SIP Intermediary | Bob | LS
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Request w/Location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Response</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;-----------------</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Figure 1. Location Conveyed by Value

In Figure 1, Alice is both the Target and the LS that is conveying her location directly to Bob, who acts as an LR. This conveyance is point-to-point - it does not pass through any SIP-layer intermediary. A Location Object appears by-value in the initial SIP request as a MIME body, and Bob responds to that SIP request as appropriate. There is a 'Bad Location Information' response code introduced within this document to specifically inform Alice if she conveys bad location to Bob (e.g., Bob "cannot parse the location provided", or "there is not enough location information to determine where Alice is").

3.2 Location Conveyed as a Location URI

Here we make Figure 1 a little more complicated by showing a diagram of indirect Location Conveyance from Alice to Bob, where Bob’s entity has to retrieve the location object from a 3rd party server.
In Figure 2, location is conveyed indirectly, via a Location URI carried in the SIP message (more of those details later). If Alice sends Bob this Location URI, Bob will need to dereference the URI — analogous to Content Indirection [RFC4483] — in order to request the location information. In general, the LS provides the location value to Bob instead of Alice directly. From a user interface perspective, Bob the user won’t know that this information was gathered from an LS indirectly rather than culled from the SIP request, and practically this does not impact the operation of location-based applications.

3.3 Location Conveyed though a SIP Intermediary

In Figure 3, we introduce the idea of a SIP intermediary into the example to illustrate the role of proxying in the location architecture. This intermediary could be a SIP proxy or it could be a back-to-back-user-agent (B2BUA). In this message flow, the SIP intermediary could act as a LR, in addition to Bob. The primary use case for intermediaries consuming location information is location-based routing. In this case, the intermediary chooses a next hop for the SIP request by consulting a specialized location service which selects forwarding destinations based on geographical location. In this case, the intermediary acts as a Location Recipient.
However, the most common case will be one in which the SIP intermediary receives a request with location information (conveyed either by-value or by-reference) and does not know or care about Alice’s location, or support this extension, and merely passes it on to Bob. In this case, the intermediary does not act as a Location Recipient.

Note that an intermediary does not have to perform location-based routing in order to be location recipient. It could be the case that a SIP intermediary which does not perform location-based routing but does care when Alice includes her location; for example, it could care that the location information is complete or that it correctly identifies where Alice is. The best example of this is intermediaries that verify location information for emergency calling, but it could also be for any location based routing - e.g., contacting Pizza Hut, making sure that organization has Alice’s proper location in the initial SIP request.

There is another scenario in which the SIP intermediary cares about location and is not an LR, one in which the intermediary inserts another location of the Target, Alice in this case, into the request, and forwards it. This secondary insertion is generally not advisable because downstream SIP entities will not be given any guidance about which location to believe is better, more reliable, less prone to error, more granular, worse than the other location or just plain wrong.

The only conceivable way forward, when a second location is placed into the same SIP request by a SIP intermediary is to take a "you break it, you bought it" philosophy with respect to the inserting SIP intermediary. That entity becomes completely responsible for all location within that SIP request (more on this in Section 4).

3.4 SIP Intermediary Replacing Bad Location

If the SIP intermediary rejects the message due to unsuitable location information (we are not going to discuss any other reasons in this document, and there are many), the SIP response will
indicate there was ‘Bad Location Information’ in the SIP request, and provide a location specific error code indicating what Alice needs to do to send an acceptable request (see Figure 4 for this scenario).

<table>
<thead>
<tr>
<th>Alice</th>
<th>SIP Intermediary</th>
<th>Bob</th>
<th>LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Request</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/Location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------</td>
<td>-------------------</td>
<td>----</td>
</tr>
<tr>
<td>Rejected</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/New Location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Request</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/New Location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Request</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/New Location</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. SIP Intermediary Replacing Bad Location

In this last use case, the SIP intermediary wishes to include a Location Object indicating where it understands Alice to be. Thus, it must inform her user agent what location she should include in any subsequent SIP request that contains her location. In these cases, the intermediary can reject Alice’s request, through the SIP response, convey to her the best way to repair the request in order for the intermediary to accept it.

Overriding location information provided by the user requires a deployment where an intermediary necessarily knows better than an end user – after all, it could be that Alice has an on-board GPS, and the SIP intermediary only knows her nearest cell tower. Which is more accurate location information? Currently, there is no way to tell which entity is more accurate, or which is wrong – for that matter. This document will not specify how to indicate which location is more accurate than another. If desired, intermediaries may furthermore allow both Alice’s internally generated location, as well as the SIP intermediary’s determination of where Alice, to appear in the same SIP request (the resubmitted one), and permit that to be forwarded to Bob. This case is discussed in more detail in section 4.2 of this document.

As an aside, it is not envisioned that any SIP-based emergency services request (i.e., IP-911, or 112 type of call attempt) will receive a corrective ‘Bad Location Information’ response from an intermediary. Most likely, the SIP intermediary would in that scenario act a B2BUA and insert into the request by-value any
appropriate location information for the benefit of Public Safety Answering Point (PSAP) call centers to expedite call reception by the emergency services personnel; thereby, minimizing any delay in call establishment time. The implementation of these specialized deployments is, however, outside the scope of this document.

4. SIP Modifications for Geolocation Conveyance

The following sections detail the modifications to SIP for location conveyance.

4.1 The Geolocation Header

This document defines "Geolocation" as a new SIP header field registered by IANA, with the following ABNF [RFC5234]:

Geolocation-header = "Geolocation" HCOLON Geolocation-value
Geolocation-value = ( locationValue [ COMMA locationValue ] ) / routing-param
locationValue = LAQUOT locationURI RAQUOT *(SEMI geoloc-param)
locationURI = sip-URI / sips-URI / pres-URI / http-URI / HTTPS-URI / cid-url ; (from RFC 2392) / absoluteURI ; (from RFC 3261)
geoloc-param = generic-param; (from RFC 3261)
routing-param = "routing-allowed" EQUAL "yes" / "no"

sip-URI, sips-URI and absoluteURI are defined according to [RFC3261].

The pres-URI is defined in [RFC3859].

HTTP-URI and HTTPS-URI are defined according to [RFC2616] and [RFC2818], respectively.

The cid-url is defined in [RFC2392] to locate message body parts. This URI type is present in a SIP request when location is conveyed as a MIME body in the SIP message.

GEO-URIs [RFC5870] are not appropriate for usage the SIP Geolocation header.

Other URI schemas used in the location URI MUST be reviewed against the RFC 3693 [RFC3693] criteria for a Using Protocol.

The Geolocation header field has zero, one or two locationValues, but MUST NOT contain more than two locationValue. A SIP intermediary SHOULD NOT add location to a SIP request that already contains location. This will quite often lead to confusion within LRs. However, if a SIP intermediary were to add location, even if location was not previously present in a SIP request, that SIP
intermediary is fully responsible for addressing the concerns of any 424 (Bad Location Information) SIP response it receives about this location addition, and MUST NOT pass on (upstream) the 424 response.

The placement of the "routing-allowed" header field parameter, strongly encouraged by [RFC5606], is outside the locationValue, and MUST always be last in the header field value. The routing-allowed parameter MUST be present, even when no locationValue is present. This scenario sets the routing-allowed policy downstream along the request’s signaling path. This header field parameter only has the values "=yes" or "=no". When this parameter is "=yes", the locationValue can be used for routing decisions along the downstream signaling path by intermediaries. If no routing-allowed parameter is present in a SIP request, a SIP intermediary MAY insert this value with a RECOMMENDED value of "no" by default.

When this parameter is "=no", this means no locationValue (inserted by the originating UAC or any intermediary along the signaling path can be used by any SIP intermediary to make routing decisions. Intermediaries that attempt to use the location information for routing purposes in spite of this counter indication may end up routing the request improperly as a result. Sections 4.3 describes the details on what a routing intermediary does if it determines it needs to use the location in the SIP request in order to process the message further.

The practical implication is that when the "routing-allowed" parameter is set to "no", if a cid:url is present in the SIP request, intermediaries MUST NOT view the location (because it is not for intermediaries to view), and if a location URI is present, intermediaries MUST NOT dereference it. UAs are allowed to view location in the SIP request even when the "routing-allowed" parameter is set to "no". An LR MUST by default consider the "routing-allowed" header parameter as set to "no", with no exceptions, unless the header field value is set to "yes".

If a routing-allowed parameter is parsed as set to "=yes", an implementation MUST parse the rest of the SIP headers for another instance of the Geolocation header value to determine if there is another instance of the routing-allowed parameter set to "=no". If this is the case, the behavior MUST be to process the "=no" indication only, and ignore the "=yes".

This document defines the Geolocation header field as valid in the following SIP requests:

- INVITE [RFC3261],
- REGISTER [RFC3261],
- OPTIONS [RFC3261],
- BYE [RFC3261],
- UPDATE [RFC3311],
- INFO [RFC2976],
- MESSAGE [RFC3428],
- REFER [RFC3515],
- SUBSCRIBE [RFC3265],
- NOTIFY [RFC3265],
- PUBLISH [RFC3903],
- PRACK [RFC3262]
The following table extends the values in Tables 2 and 3 of RFC 3261 [RFC3261].

<table>
<thead>
<tr>
<th>Header field</th>
<th>where proxy</th>
<th>INV</th>
<th>ACK</th>
<th>CAN</th>
<th>BYE</th>
<th>REG</th>
<th>OPT</th>
<th>PRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geolocation</td>
<td>R ar o - - o o o o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geolocation</td>
<td>424 r o - - o o o o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Header field</th>
<th>where proxy</th>
<th>SUB</th>
<th>NOT</th>
<th>UPD</th>
<th>MSG</th>
<th>REF</th>
<th>INF</th>
<th>PUB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geolocation</td>
<td>R ar o o o o o o o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geolocation</td>
<td>424 r o o o o o o o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Summary of the Geolocation Header Field

The Geolocation header field MAY be included in any one of the optional requests by a UA. A proxy MAY add the Geolocation header field, but MUST NOT modify any pre-existing locationValue, including the "routing-allowed" header field value.

A SIP intermediary MAY add a Geolocation header field if one is not present - for example, when a user agent does not support the Geolocation mechanism but their outbound proxy does and knows their location, or any of a number of other use cases (see Section 3). When adding a Geolocation header, a SIP intermediary MAY supply the "routing-allowed" parameter if not yet present in the SIP request, but MUST NOT add a "routing-allowed" parameter if one is already present in this SIP request.

SIP implementations are advised to pay special attention to the policy elements for location retransmission and retention described in RFC 4119.

4.2 424 (Bad Location Information) Response Code

This SIP extension creates a new location-specific response code, defined as follows,

    424 (Bad Location Information)

The 424 (Bad Location Information) response code is a rejection of the request due to its location contents, indicating location information that was malformed or not satisfactory for the recipient’s purpose, or could not be dereferenced.

A SIP intermediary can also reject a location it receives from a Target when it understands the Target to be in a different location. The proper handling of this scenario, described in Section 3.4, is for the SIP intermediary to include the proper location in the 424 Response. This SHOULD be included in the response as a MIME message.
body (i.e., a location value), rather than as a URI; however, in cases where the intermediary is willing to share location with recipients but not with a user agent, a reference might be necessary.

As mentioned in Section 3.4, it might be the case that the intermediary does not want to chance providing less accurate location information than the user agent; thus it will compose its understanding of where the user agent is in a separate <geopriv> element of the same PIDF-LO message body in the SIP response (which also contains the Target’s version of where it is). Therefore, both locations are included - each with different <method> elements. The proper reaction of the user agent is to generate a new SIP request that includes this composed location object, and send it towards the original LR. SIP intermediaries can verify that subsequent requests properly insert the suggested location information before forwarding said requests.

SIP intermediaries MUST NOT add, modify or delete the location in a 424 response. This specifically applies to intermediaries that are between the 424 response generator and the original UAC. All respects of the Geolocation and Geolocation-Error headers and PIDF-LO(s) MUST remain unchanged, never added to or deleted.

Section 4.3 describes a Geolocation-Error header field to provide more detail about what was wrong with the location information in the request. This header field MUST be included in the 424 response.

It is only appropriate to generate a 424 response when the responding entity needs a locationValue and there are no locationValues included in the SIP request that are usable by that recipient, or as shown in Figure 4 of section 3.4. In this scenario, a SIP intermediary is informing the upstream UA which location to include in the next SIP request.

A 424 MUST NOT be sent in response to a request that lacks a Geolocation header entirely, as the user agent in that case may not support this extension at all. If a SIP intermediary inserted a locationValue into a SIP request where one was not previously present, it MUST take any and all responsibility for the corrective action if it receives a 424 to a SIP request it sent.

A 424 (Bad Location Information) response is a final response within a transaction, and MUST NOT terminate an existing dialog.

4.3 The Geolocation-Error Header

As discussed in Section 4.2, more granular error notifications specific to location errors within a received request are required if the UA is to know what was wrong within the original request. The Geolocation-Error header field is used for this purpose.
The Geolocation-Error header field is used to convey location-specific errors within a response. The Geolocation-Error header field has the following ABNF [RFC5234]:

Geolocation-Error = "Geolocation-Error" HCOLON
locationErrorValue

locationErrorValue = location-error-arg
location-error-arg  = location-error-code
 *(SEMI location-error-params)
location-error-code = 1*3DIGIT
location-error-params = location-error-code-text
/generic-param ; from RFC3261
location-error-code-text = "code" EQUAL quoted-string ; from RFC3261

The Geolocation-Error header field MUST contain only one locationErrorValue to indicate what was wrong with the locationValue the Location Recipient determined was bad. The locationErrorValue contains a 3-digit error code indicating what was wrong with the location in the request. Each error code has a corresponding quoted error text string that is human understandable. This text string is OPTIONAL, but RECOMMENDED for human readability.

The following table extends the values in Table 2&3 of RFC 3261 [RFC3261].

<table>
<thead>
<tr>
<th>Header field</th>
<th>where proxy</th>
<th>INV</th>
<th>ACK</th>
<th>CAN</th>
<th>BYE</th>
<th>REG</th>
<th>OPT</th>
<th>PRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geolocation-Error</td>
<td>r</td>
<td>ar</td>
<td>o</td>
<td>-</td>
<td>-</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Header field</th>
<th>where proxy</th>
<th>SUB</th>
<th>NOT</th>
<th>UPD</th>
<th>MSG</th>
<th>REF</th>
<th>INF</th>
<th>PUB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geolocation-Error</td>
<td>r</td>
<td>ar</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
</tbody>
</table>

Table 2: Summary of the Geolocation-Error Header Field

The Geolocation-Error header field MAY be included in any response to one of the above SIP requests, so long as a Geolocation locationValue was in the request part of the transaction. For example, Alice includes her location in an INVITE to Bob. Bob can accept this INVITE, thus creating a dialog, even though his UA determined the location contained in the INVITE was bad. Bob merely includes a Geolocation-Error header value in the 200 OK to the INVITE informing Alice the INVITE was accepted but the location provided was bad. The SIP requests included in table 2 above are the ones allowed to optionally contain the Geolocation header field (see section 4.1).

If, on the other hand, Bob cannot accept Alice’s INVITE without a suitable location, a 424 (Bad Location Information) is sent. This
message flow is shown in Figures 1, 2 or 3 in Section 3.

The following subsections provide an initial list of location based errors for any SIP non-100 response, including the new 424 (Bad Location Information) response. These error codes are divided into 4 categories, based on how the response receiver should react to these errors.

- 1XX errors mean the LR cannot process the location within the request.
- 2XX errors mean the LR wants the LS to send new or updated location information, perhaps with a delay associated with when to send the request.
- 3XX errors mean some specific permission is necessary to process the included location information.
- 4XX errors mean there was trouble dereferencing the Location URI sent.

Within these 4 number ranges, there is a top level error as follows:

Geolocation-Error: 100 "Cannot Process Location"
Geolocation-Error: 200 "Retry Location Later with device updated location"
Geolocation-Error: 300 "Permission To Use Location Information"
Geolocation-Error: 400 "Dereference Failure"

There are two specific Geolocation-Error codes necessary to include in this document, both have to do with permissions necessary to process the SIP request; they are

Geolocation-Error: 301 "Permission To Retransmit Location Information to a Third Party"

This location error is specific to having the Presence Information Data Format (PIDF-LO) [RFC4119] <retransmission-allowed> element set to "=no". This location error is stating it requires permission (i.e., PIDF-LO <retransmission-allowed> element set to "=yes") to process this SIP request further. If the LS sending the location information does not want to give this permission, it will not reset this permission in a new request. If the LS wants this message processed without this permission reset, it MUST choose another logical path (if one exists).

Geolocation-Error: 302 "Permission to Route based on Location Information"
This location error is specific to having the locationValue header parameter <routing-allowed> set to "=no". This location error is stating it requires permission (i.e., a <routing-allowed> set to "=yes") to process this SIP request further. If the LS sending the location information does not want to give this permission, it will not reset this permission in a new request. If the LS wants this message processed without this permission reset, it MUST choose another logical path (if one exists).

4.4 The 'geolocation' Option Tag

This document creates and registers with the IANA one new option tag: "geolocation". This option tag is to be used, as defined in [RFC3261], in the Require, Supported and Unsupported header fields.

4.5 Location URIs in Message Bodies

In the case where a location recipient sends a 424 response and wishes to communicate suitable location by reference rather than by value, the 424 MUST include a content-indirection body per RFC 4483.

4.6 Location URIs Allowed

The following is part of the discussion started in Section 3, Figure 2, which introduced the concept of sending location indirectly.

If a location URI is included in a SIP request, it SHOULD be a SIP-, SIPS- or PRES-URI. When PRES: is used, as defined in [RFC3856], if the resulting resolution resolves to a SIP: or SIPS: URI, this section applies. These schemes MUST be implemented according to this document.

See [ID-GEO-FILTERS] for more details on dereferencing location.

An HTTP: [RFC2616] or HTTPS: URI [RFC2818] are also allowed, and SHOULD be dereferenced according to [ID-HELD-DEREF].

5. Geolocation Examples

5.1 Location-by-value (in Coordinate Format)

This example shows an INVITE message with a coordinate location. In this example, the SIP request uses a sips-URI [RFC3261], meaning this message is protected using TLS on a hop-by-hop basis.

INVITE sips:bob@biloxi.example.com SIP/2.0
Via: SIPS/2.0/TLS pc33.atlanta.example.com;branch=z9hG4bK74bf9
Max-Forwards: 70
To: Bob <sips:bob@biloxi.example.com>
From: Alice <sips:alice@atlanta.example.com>;tag=9fxced76sl
Call-ID: 3848276298220188511@atlanta.example.com
Geolocation: <cid:target123@atlanta.example.com>
    ;routing-allowed=no
Supported: geolocation
Accept: application/sdp, application/pidf+xml
CSeq: 31862 INVITE
Contact: <sips:alice@atlanta.example.com>
Content-Type: multipart/mixed; boundary=boundary1
Content-Length: ...

--boundary1
Content-Type: application/sdp

...SDP goes here

--boundary1
Content-Type: application/pidf+xml
Content-ID: <target123@atlanta.example.com>
<?xml version="1.0" encoding="UTF-8"?>
<presence
    xmlns="urn:ietf:params:xml:ns:pidf"
    xmlns:gp="urn:ietf:params:xml:ns:pidf:geopriv10"
    xmlns:gml="http://www.opengis.net/gml"
    entity="pres:alice@atlanta.example.com">
<dm:device id="target123-1">
<gp:geopriv>
<gp:location-info>
<gml:location>
    <gml:Point srsName="urn:ogc:def:crs:EPSG::4326">
        <gml:pos>32.86726 -97.16054</gml:pos>
    </gml:Point>
</gml:location>
</gp:location-info>
<gp:usage-rules>
<gbp:retransmission-allowed>false
<gbp:retransmission-allowed>
<gbp:retention-expiry>2010-11-14T20:00:00Z
</gbp:retention-expiry>
</gp:usage-rules>
<gp:method>802.11</gp:method>
</gp:geopriv>
</dm:device>
</presence>

--boundary1--
The Geolocation header field from the above INVITE:

Geolocation: <cid:target123@atlanta.example.com>

... indicates the content-ID location [RFC2392] within the multipart message body of where location information is. An assumption can be made that SDP is the other message body part. The "cid:" eases message body parsing by disambiguating the MIME body that contains the location information associated with this request.

If the Geolocation header field did not contain a "cid:" scheme, for example, it could look like this location URI:

Geolocation: <sips:target123@server5.atlanta.example.com>

... the existence of a non-"cid:" scheme indicates this is a location URI, to be dereferenced to learn the Target’s location. Any node wanting to know where user "target123" is would subscribe to that user at server5 to dereference the sips-URI (see Figure 3 in section 3 for this message flow).

5.2 Two Locations Composed in Same Location Object Example

This example shows the INVITE message after a SIP intermediary rejected the original INVITE (say, the one in section 5.1). This INVITE contains the composed LO sent by the SIP intermediary which includes where the intermediary understands Alice to be. The rules of RFC 5491 [RFC5491] are followed in this construction.

This example is here, but should not be taken as occurring very often. In fact, this is believed to be a corner case of location conveyance applicability.

INVITE sips:bob@biloxi.example.com SIP/2.0
Via: SIPS/2.0/TLS pc33.atlanta.example.com;branch=z9hG4bK74bf0
Max-Forwards: 70
To: Bob <sips:bob@biloxi.example.com>
From: Alice <sips:alice@atlanta.example.com>;tag=9fxced76sl
Call-ID: 3848276298220188512@atlanta.example.com
Geolocation: <cid:target123@atlanta.example.com>
    ;routing-allowed=no
Supported: geolocation
Accept: application/sdp, application/pidf+xml
CSeq: 31863 INVITE
Contact: <sips:alice@atlanta.example.com>
Content-Type: multipart/mixed; boundary=boundary1
Content-Length: ...

--boundary1

Content-Type: application/sdp
Content-Type: application/pidf+xml
Content-ID: <target123@atlanta.example.com>

<?xml version="1.0" encoding="UTF-8"?>
<pres>
  <dm:
device id="target123-1">
    <gp:geopriv>
      <gp:location-info>
        <gml:location>
          <gml:Point srsName="urn:ogc:def:crs:EPSG::4326">
            <gml:pos>32.86726 -97.16054</gml:pos>
          </gml:Point>
        </gml:location>
        <gp:usage-rules>
          <gbp:retransmission-allowed>false</gbp:retransmission-allowed>
          <gbp:retention-expiry>2010-11-14T20:00:00Z</gbp:retention-expiry>
        </gp:usage-rules>
        <gp:method>802.11</gp:method>
      </gp:location-info>
      <dm:deviceID>mac:1234567890ab</dm:deviceID>
      <dm:timestamp>2010-11-04T20:57:29Z</dm:timestamp>
    </gp:geopriv>
  </dm:device>
  <dm:person id="target123">
    <gp:geopriv>
      <gp:location-info>
        <cl:civicAddress>
          <cl:country>US</cl:country>
          <cl:A1>Texas</cl:A1>
          <cl:A3>Colleyville</cl:A3>
          <cl:RD>Treemont</cl:RD>
          <cl:STS>Circle</cl:STS>
          <cl:HNO>3913</cl:HNO>
          <cl:FLR>1</cl:FLR>
          <cl:NAM>Haley’s Place</cl:NAM>
          <cl:PC>76034</cl:PC>
        </cl:civicAddress>
      </gp:location-info>
    </gp:geopriv>
  </dm:person>
</pres>
6. Geopriv Privacy Considerations

Location information is considered by most to be highly sensitive information, requiring protection from eavesdropping and altering in transit. [RFC3693] originally articulated rules to be followed by any protocol wishing to be considered a "Using Protocol", specifying how a transport protocol meets those rules. [ID-GEOPRIV-ARCH] updates the guidance in RFC3693 to include subsequently-introduced entities and concepts in the geolocation architecture. Implementations of this SIP location conveyance mechanism MUST adhere to the guidance given in RFC3693 and its successors, including (but not limited to) the handling of rules for retention and retransmission.

7. Security Considerations

Conveyance of physical location of a UA raises privacy concerns, and depending on use, there probably will be authentication and integrity concerns. This document calls for conveyance to be accomplished through secure mechanisms, like S/MIME encrypting message bodies (although this is not widely deployed), TLS protecting the overall signaling or conveyance location by-reference and requiring all entities that dereference location to authenticate themselves. In location-based routing cases, encrypting the location payload with an end-to-end mechanism such as S/MIME is problematic, because one or more proxies on the path need the ability to read the location information to retarget the message to the appropriate new destination UAS. Data can only be encrypted to a particular, anticipated target, and thus if multiple recipients need to inspect a piece of data, and those recipients cannot be predicted by the sender of data, encryption is not a very feasible choice. Securing the location hop-by-hop, using TLS, protects the message from eavesdropping and modification in transit, but exposes the information to all proxies on the path as well as the endpoint. In most cases, the UA has no trust relationship with the proxy or proxies providing location-based routing services, so such end-to-middle solutions might not be appropriate either.
When location information is conveyed by reference, however, one can properly authenticate and authorize each entity that wishes to inspect location information. This does not require that the sender of data anticipate who will receive data, and it does permit multiple entities to receive it securely, but it does not however obviate the need for pre-association between the sender of data and any prospective recipients. Obviously, in some contexts this pre-association cannot be presumed; when it is not, effectively unauthenticated access to location information must be permitted. In this case, choosing pseudo-random URIs for location by-reference, coupled with path encryption like SIPS, can help to ensure that only entities on the SIP signaling path learn the URI, and thus restores rough parity with sending location by-value.

Location information is especially sensitive when the identity of its Target is obvious. Note that there is the ability, according to [RFC3693] to have an anonymous identity for the Target’s location. This is accomplished by use of an unlinkable pseudonym in the "entity=" attribute of the <presence> element [RFC4479]. Though, this can be problematic for routing messages based on location (covered in the document above). Moreover, anyone fishing for information would correlate the identity at the SIP layer with that of the location information referenced by SIP signaling.

When a UA inserts location, the UA sets the policy on whether to reveal its location along the signaling path - as discussed in Section 4, as well as flags in the PIDF-LO [RFC4119]. UAC implementations MUST make such capabilities conditional on explicit user permission, and MUST alert the user that location is being conveyed.

This SIP extension offers the default ability to require permission to view location while the SIP request is in transit. The default for this is set to "no". There is an error explicitly describing how an intermediary asks for permission to view the Target’s location, plus a rule stating the user has to be made aware of this permission request.

There is no end-to-end integrity on any locationValue or locationErrorValue header field parameter (or middle-to-end if the value was inserted by a intermediary), so recipients of either header field need to implicitly trust the header field contents, and take whatever precautions each entity deems appropriate given this situation.

8. IANA Considerations

The following are the IANA considerations made by this SIP extension. Modifications and additions to these registrations require a standards track RFC (Standards Action).

[Editor’s Note: RFC-Editor – within the IANA section, please
8.1 IANA Registration for the SIP Geolocation Header Field

The SIP Geolocation Header Field is created by this document, with its definition and rules in Section 4.1 of this document, and should be added to the IANA sip-parameters registry, in the portion titled "Header Field Parameters and Parameter Values".

<table>
<thead>
<tr>
<th>Header Field</th>
<th>Parameter Name</th>
<th>Values</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geolocation</td>
<td>routing-allowed</td>
<td>yes</td>
<td>[this doc]</td>
</tr>
</tbody>
</table>

8.2 IANA Registration for New SIP 'geolocation' Option Tag

The SIP option tag "geolocation" is created by this document, with the definition and rule in Section 4.4 of this document, to be added to the IANA sip-parameters registry.

8.3 IANA Registration for 424 Response Code

Reference: RFC-XXXX (i.e., this document)
Response code: 424 (recommended number to assign)
Default reason phrase: Bad Location Information

This SIP Response code is defined in section 4.2 of this document.

8.4 IANA Registration of New Geolocation-Error Header Field

The SIP Geolocation-error header field is created by this document, with its definition and rules in Section 4.3 of this document, to be added to the IANA sip-parameters registry, in the portion titled "Header Field Parameters and Parameter Values".

<table>
<thead>
<tr>
<th>Header Field</th>
<th>Parameter Name</th>
<th>Values</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geolocation-Error</td>
<td>code=</td>
<td>yes*</td>
<td>[this doc]</td>
</tr>
</tbody>
</table>

* see section 8.5 for the newly created values.

8.5 IANA Registration for the SIP Geolocation-Error Codes

New location specific Geolocation-Error codes are created by this document, and registered in a new table in the IANA sip-parameters
registry. Details of these error codes are in Section 4.3 of this document.

Geolocation-Error codes
-----------------------
Geolocation-Error codes provide reason for the error discovered by Location Recipients, categorized by action to be taken by error recipient.

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>&quot;Cannot Process Location&quot;</td>
<td>[this doc]</td>
</tr>
<tr>
<td>200</td>
<td>&quot;Retry Location Later with device updated location&quot;</td>
<td>[this doc]</td>
</tr>
<tr>
<td>300</td>
<td>&quot;Permission To Use Location Information&quot;</td>
<td>[this doc]</td>
</tr>
<tr>
<td>301</td>
<td>&quot;Permission To Retransmit Location Information to a Third Party&quot;</td>
<td>[this doc]</td>
</tr>
<tr>
<td>302</td>
<td>&quot;Permission to Route based on Location Information&quot;</td>
<td>[this doc]</td>
</tr>
<tr>
<td>400</td>
<td>&quot;Dereference Failure&quot;</td>
<td>[this doc]</td>
</tr>
</tbody>
</table>

8.6 IANA Registration of Location URI Schemes

This document directs IANA to create a new set of parameters in a separate location from SIP and Geopriv, called the "Location Reference URI" registry, containing the URI scheme, the Content-Type, and the reference, as follows:

<table>
<thead>
<tr>
<th>URI Scheme</th>
<th>Content-Type</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIP:</td>
<td></td>
<td>[this doc]</td>
</tr>
<tr>
<td>SIPS:</td>
<td></td>
<td>[this doc]</td>
</tr>
<tr>
<td>PRES:</td>
<td></td>
<td>[this doc]</td>
</tr>
<tr>
<td>HTTP:</td>
<td></td>
<td>[this doc]</td>
</tr>
<tr>
<td>HTTPS:</td>
<td></td>
<td>[this doc]</td>
</tr>
</tbody>
</table>

Additions to this registry must be defined in a permanent and readily available specification (this is the "Specification Required" IANA policy defined in [RFC5226]).

9. Acknowledgements

To Dave Oran for helping to shape this idea.
To Dean Willis for guidance of the effort.
To Allison Mankin, Dick Knight, Hannes Tschofenig, Henning Polk, et al.
Schulzrinne, James Winterbottom, Jeroen van Bemmel, Jean-Francois Mule, Jonathan Rosenberg, Keith Drage, Marc Linsner, Martin Thomson, Mike Hammer, Ted Hardie, Shida Shubert, Umesh Sharma, Richard Barnes, Dan Wing, Matt Lepinski, John Elwell and Jacqueline Lee for constructive feedback and nits checking.

Special thanks to Paul Kyzivat for his help with the ABNF in this document and to Robert Sparks for many helpful comments and the proper construction of the Geolocation-Error header field.

And finally, to Spencer Dawkins for giving this doc a good scrubbing to make it more readable.

10. References

10.1 Normative References


10.2 Informative References


Appendix A. Requirements for SIP Location Conveyance

The following subsections address the requirements placed on the UAC, the UAS, as well as SIP proxies when conveying location. If a requirement is not obvious in intent, a motivational statement is included below it.

A.1 Requirements for a UAC Conveying Location

UAC-1  The SIP INVITE Method [RFC3261] must support location conveyance.

UAC-2  The SIP MESSAGE method [RFC3428] must support location conveyance.
UAC-3 SIP Requests within a dialog should support location conveyance.

UAC-4 Other SIP Requests may support location conveyance.

UAC-5 There must be one, mandatory to implement means of transmitting location confidentially.

Motivation: to guarantee interoperability.

UAC-6 It must be possible for a UAC to update location conveyed at any time in a dialog, including during dialog establishment.

Motivation: if a UAC has moved prior to the establishment of a dialog between UAs, the UAC must be able to send location information. If location has been conveyed, and the UA moves, the UAC must be able to update the location previously conveyed to other parties.

UAC-7 The privacy and security rules established within [RFC3693] that would categorize SIP as a ‘Using Protocol’ must be met.

UAC-8 The PIDF-LO [RFC4119] is a mandatory to implement format for location conveyance within SIP.

Motivation: interoperability with other IETF location protocols and Mechanisms.

UAC-9 There must be a mechanism for the UAC to request the UAS send its location.

UAC-9 has been DEPRECATED by the SIP WG, due to the many problems this requirement would have caused if implemented. The solution is for the above UAS to send a new request to the original UAC with the UAS’s location.

UAC-10 There must be a mechanism to differentiate the ability of the UAC to convey location from the UACs lack of knowledge of its location.

Motivation: Failure to receive location when it is expected can happen because the UAC does not implement this extension, or because the UAC implements the extension, but does not know where the Target is. This may be, for example, due to the failure of the access network to provide a location acquisition mechanism the UAC supports. These cases must be differentiated.

UAC-11 It must be possible to convey location to proxy servers
along the path.

Motivation: Location-based routing.

A.2 Requirements for a UAS Receiving Location

The following are the requirements for location conveyance by a UAS:

UAS-1 SIP Responses must support location conveyance.

Just as with UAC-9, UAS-1 has been DEPRECATED by the SIP WG, due to the many problems this requirement would have caused if implemented. The solution is for the above UAS to send a new request to the original UAC with the UAS’s location.

UAS-2 There must be a unique 4XX response informing the UAC it did not provide applicable location information.

In addition, requirements UAC-5, 6, 7 and 8 also apply to the UAS.

A.3 Requirements for SIP Proxies and Intermediaries

The following are the requirements for location conveyance by a SIP proxies and intermediaries:

Proxy-1 Proxy servers must be capable of adding a Location header field during processing of SIP requests.

Motivation: Provide network assertion of location when UACs are unable to do so, or when network assertion is more reliable than UAC assertion of location

Note: Because UACs connected to SIP signaling networks may have widely varying access network arrangements, including VPN tunnels and roaming mechanisms, it may be difficult for a network to reliably know the location of the endpoint. Proxy assertion of location is NOT RECOMMENDED unless the SIP signaling network has reliable knowledge of the actual location of the Targets.

Proxy-2 There must be a unique 4XX response informing the UAC it did not provide applicable location information.
Re-INVITE and Target-refresh Request Handling in the Session Initiation Protocol (SIP)
draft-ietf-sipcore-reinvite-08

Abstract

The procedures for handling SIP re-INVITEs are described in RFC 3261. Implementation and deployment experience has uncovered a number of issues with the original documentation, and this document provides additional procedures that update the original specification to address those issues. In particular, this document defines in which situations a UAS (User Agent Server) should generate a success response and in which situations a UAS should generate an error response to a re-INVITE. Additionally, this document defines further details of procedures related to target-refresh requests.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on June 22, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents
Table of Contents

1. Introduction ................................................. 3
2. Terminology ................................................ 3
3. Changing the Session State During a Re-INVITE .................. 4
   3.1. Background on Re-INVITE Handling by UASs ................. 4
   3.2. Problems with Error Responses and Already-executed Changes ................................................. 8
   3.3. UAS Behavior ............................................ 9
   3.4. UAC Behavior ............................................ 10
   3.5. Glare Situations ....................................... 11
   3.6. Example of UAS Behavior ................................ 11
   3.7. Example of UAC Behavior ................................ 14
   3.8. Clarifications on Cancelling Re-INVITEs .................. 16
4. Refreshing a Dialog’s Targets .................................. 16
   4.1. Background and Terminology on a Dialog’s Targets ........ 16
   4.2. Background on Target-refresh Requests ..................... 17
   4.3. Clarification on the Atomicity of Target-Refresh Requests ......................................................... 17
   4.4. UA Updating the Dialog’s Local Target in a Request ..... 18
   4.5. UA Updating the Dialog’s Local Target in a Response ... 18
   4.6. A Request Updating the Dialog’s Remote Target .......... 19
   4.7. A Response Updating the Dialog’s Remote Target ......... 19
   4.8. Race Conditions and Target Refreshes ................... 20
   4.9. Early Dialogs ............................................ 20
5. A UA Losing its Contact ......................................... 21
   5.1. Background on re-INVITE Transaction Routing ............ 21
   5.2. Problems with UAs Losing their Contact ................. 21
   5.3. UAS Losing its Contact: UAC Behavior .................... 22
   5.4. UAC Losing its Contact: UAS Behavior .................... 22
   5.5. UAC Losing its Contact: UAC Behavior .................... 23
6. Security Considerations ........................................ 23
7. IANA Considerations ........................................... 24
8. Acknowledgements .............................................. 24
9. References .................................................... 24
   9.1. Normative References .................................... 24
   9.2. Informative References ................................... 24
Authors’ Addresses ................................................ 25
1. Introduction

As discussed in Section 14 of RFC 3261 [RFC3261], an INVITE request sent within an existing dialog is known as a re-INVITE. A re-INVITE is used to modify session parameters, dialog parameters, or both. That is, a single re-INVITE can change both the parameters of its associated session (e.g., changing the IP address where a media stream is received) and the parameters of its associated dialog (e.g., changing the remote target of the dialog). A re-INVITE can change the remote target of a dialog because it is a target refresh request, as defined in Section 6 of RFC 3261 [RFC3261].

A re-INVITE transaction has an offer/answer [RFC3264] exchange associated to it. The UAC (User Agent Client) generating a given re-INVITE can act as the offerer or as the answerer. A UAC willing to act as the offerer includes an offer in the re-INVITE. The UAS (User Agent Server) then provides an answer in a response to the re-INVITE. A UAC willing to act as answerer does not include an offer in the re-INVITE. The UAS then provides an offer in a response to the re-INVITE becoming, thus, the offerer.

Certain transactions within a re-INVITE (e.g., UPDATE [RFC3311] transactions) can also have offer/answer exchanges associated to them. A UA (User Agent) can act as the offerer or the answerer in any of these transactions regardless of whether the UA was the offerer or the answerer in the umbrella re-INVITE transaction.

There has been some confusion among implementors regarding how a UAS should handle re-INVITEs. In particular, implementors requested clarification on which type of response a UAS should generate in different situations. In this document, we clarify these issues.

Additionally, there has also been some confusion among implementors regarding target refresh requests, which include but are not limited to re-INVITEs. In this document, we also clarify the process by which remote targets are refreshed.

Indented passages such as this one are used in this document to provide additional information and clarifying text. They do not contain normative protocol behavior.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].
3. Changing the Session State During a Re-INVITE

The following sections discuss how to change the state of the session during a re-INVITE transaction.

3.1. Background on Re-INVITE Handling by UASs

A UAS receiving a re-INVITE will need to, eventually, generate a response to it. Some re-INVITEs can be responded to immediately because their handling does not require user interaction (e.g., changing the IP address where a media stream is received). The handling of other re-INVITEs requires user interaction (e.g., adding a video stream to an audio-only session). Therefore, these re-INVITEs cannot be responded to immediately.

An error response to a re-INVITE has the following semantics. As specified in Section 12.2.2 of RFC 3261 [RFC3261], if a re-INVITE is rejected, no state changes are performed. These state changes include state changes associated to the re-INVITE transaction and all other transactions within the re-INVITE (this section deals with changes to the session state; target refreshes are discussed in Section 4.2). That is, the session state is the same as before the re-INVITE was received. The example in Figure 1 illustrates this point.
The UAs perform an offer/answer exchange to establish an audio-only session:

SDP1:
   m=audio 30000 RTP/AVP 0

SDP2:
   m=audio 31000 RTP/AVP 0

At a later point, the UAC sends a re-INVITE (4) in order to add a video stream to the session.

SDP3:
   m=audio 30000 RTP/AVP 0
   m=video 30002 RTP/AVP 31

The UAS is configured to automatically reject video streams. Consequently, the UAS returns an error response (5). At that point, the session parameters in use are still those resulting from the initial offer/answer exchange, which are described by SDP1 and SDP2. That is, the session state is the same as before the re-INVITE was received.

In the previous example, the UAS rejected all the changes requested in the re-INVITE by returning an error response. However, there are
situations where a UAS wants to accept some but not all the changes requested in a re-INVITE. In these cases, the UAS generates a 200 (OK) response with an SDP indicating which changes were accepted and which were not. The example in Figure 2 illustrates this point.

UAC                                           UAS

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-------------(1) INVITE SDP1-------------------</td>
<td>&lt;-------------(2) 200 OK SDP2-----------------</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>-------------(3) ACK---------------------------</td>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>-------------(4) INVITE SDP3-------------------</td>
<td>&lt;-------------(5) 200 OK SDP4-----------------</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>-------------(6) ACK---------------------------</td>
<td>----------------------------------------------</td>
</tr>
</tbody>
</table>

Figure 2: Automatic rejection of a video stream

The UAs perform an offer/answer exchange to establish an audio only session:

SDP1:
  m=audio 30000 RTP/AVP 0
  c=IN IP4 192.0.2.1

SDP2:
  m=audio 31000 RTP/AVP 0
  c=IN IP4 192.0.2.5

At a later point, the UAC moves to an access that provides a higher-bandwidth. Therefore, the UAC sends a re-INVITE (4) in order to change the IP address where it receives the audio stream to its new IP address and add a video stream to the session.
The UAS is automatically configured to reject video streams. However, the UAS needs to accept the change of the audio stream’s remote IP address. Consequently, the UAS returns a 200 (OK) response and sets the port of the video stream to zero in its SDP.

In the previous example, the UAS was configured to automatically reject the addition of video streams. The example in Figure 3 assumes that the UAS requires its user’s input in order to accept or reject the addition of a video stream and uses reliable provisional responses [RFC3262] (PRACK transactions are not shown for clarity).

UAC                                          UAS

|------------------(1) INVITE SDP1------------------>
|<--------------(2) 200 OK SDP2------------------>
|------------------(3) ACK---------------------->

|------------------(4) INVITE SDP3------------------>
|<----(5) 183 Session Progress SDP4------------->

|<--------------(6) UPDATE SDP5------------------>
|------------------(7) 200 OK SDP6------------------>
|<--------------(8) 200 OK---------------------->
|------------------(9) ACK---------------------->
Figure 3: Manual rejection of a video stream by the user

Everything up to (4) is identical to the previous example. In (5), the UAS accepts the change of the audio stream’s remote IP address but does not accept the video stream yet (it provides a null IP address instead of setting the stream to ‘inactive’ because inactive streams still need to exchange RTCP traffic).

SDP4:
  m=audio 31000 RTP/AVP 0
  c=IN IP4 192.0.2.5
  m=video 31002 RTP/AVP 31
  c=IN IP4 0.0.0.0

At a later point, the UAS’s user rejects the addition of the video stream. Consequently, the UAS sends an UPDATE request (6) setting the port of the video stream to zero in its offer.

SDP5:
  m=audio 31000 RTP/AVP 0
  c=IN IP4 192.0.2.5
  m=video 0 RTP/AVP 31
  c=IN IP4 0.0.0.0

The UAC returns a 200 (OK) response (7) to the UPDATE with the following answer:

SDP6:
  m=audio 30000 RTP/AVP 0
  c=IN IP4 192.0.2.2
  m=video 0 RTP/AVP 31

The UAS now returns a 200 (OK) response (8) to the re-INVITE.

In all the previous examples, the UAC of the re-INVITE transaction was the offerer. Examples with UACs acting as the answerers would be similar.

3.2. Problems with Error Responses and Already-executed Changes

Section 3.1 contains examples on how a UAS rejects all the changes requested in a re-INVITE without executing any of them by returning an error response (Figure 1), and how a UAS executes some of the changes requested in a re-INVITE and rejects some of them by returning a 2xx response (Figure 2 and Figure 3). A UAS can accept
and reject different sets of changes simultaneously (Figure 2) or at
different times (Figure 3).

The scenario that created confusion among implementors consists of a
UAS that receives a re-INVITE, executes some of the changes requested
in it, and then wants to reject all those already-executed changes
and revert to the pre-re-INVITE state. Such a UAS may consider
returning an error response to the re-INVITE (the message flow would
be similar to the one in Figure 1), or using an UPDATE request to
revert to the pre-re-INVITE state and then returning a 2xx response
to the re-INVITE (the message flow would be similar to the one in
Figure 3). This section explains the problems associated with
returning an error response in these circumstances. In order to
avoid these problems, the UAS should use the latter option (UPDATE
request plus a 2xx response). Section 3.3 and Section 3.4 contain
the normative statements needed to avoid these problems.

The reason for not using an error response to undo already executed
changes is that an error response to a re-INVITE for which changes
have already been executed (e.g., as a result of UPDATE transactions
or reliable provisional responses) is effectively requesting a change
in the session state. However, the UAC has no means to reject that
change if it is unable to execute them. That is, if the UAC is
unable to revert to the pre-re-INVITE state, it will not be able to
communicate this fact to the UAS.

3.3. UAS Behavior

UASs should only return an error response to a re-INVITE if no
changes to the session state have been executed since the re-INVITE
was received. Such an error response indicates that no changes have
been executed as a result of the re-INVITE or any other transaction
within it.

If any of the changes requested in a re-INVITE or in any transaction
within it have already been executed, the UAS SHOULD return a 2xx
response.

A change to the session state is considered to have been executed if
an offer/answer without preconditions [RFC4032] for the stream has
completed successfully or the UA has sent or received media using the
new parameters. Connection establishment messages (e.g., TCP SYN),
connectivity checks (e.g., when using ICE [RFC5245]), and any other
messages used in the process of meeting the preconditions for a
stream are not considered media.
Normally, a UA receiving media can easily detect when the new parameters for the media stream are used (e.g., media is received on a new port). However, in some scenarios the UA will have to process incoming media packets in order to detect whether they use the old or the new parameters.

The successful completion of an offer/answer exchange without preconditions indicates that the new parameters for the media stream are already considered to be in use. The successful completion of an offer/answer exchange with preconditions means something different. The fact that all mandatory preconditions for the stream are met indicates that the new parameters for the media stream are ready to be used. However, they will not actually be used until the UAS decides to use them. During a session establishment, the UAS can wait before using the media parameters until the callee starts being alerted or until the callee accepts the session. During a session modification, the UAS can wait until its user accepts the changes to the session. When dealing with streams where the UAS sends media more or less continuously, the UAC notices that the new parameters are in use because the UAC receives media that uses the new parameters. However, this mechanism does not work with other types of streams. Therefore, it is RECOMMENDED that when a UAS decides to start using the new parameters for a stream for which all mandatory preconditions have been met, the UAS either sends media using the new parameters or sends a new offer where the precondition-related attributes for the stream have been removed. As indicated above, the successful completion of an offer/answer exchange without preconditions indicates that the new parameters for the media stream are already considered to be in use.

3.4. UAC Behavior

A UAC that receives an error response to a re-INVITE that undoes already-executed changes within the re-INVITE may be facing a legacy UAS that does not support this specification (i.e., a UAS that does not follow the guidelines in Section 3.3). There are also certain race condition situations that get both user agents out of synchronization. In order to cope with these race condition situations, a UAC that receives an error response to a re-INVITE for which changes have been already executed SHOULD generate a new re-INVITE or UPDATE request in order to make sure that both UAs have a common view of the state of the session (the UAC uses the criteria in Section 3.3 in order to decide whether or not changes have been executed for a particular stream). The purpose of this new offer/answer exchange is to synchronize both UAs, not to request changes that the UAS may choose to reject. Therefore, session parameters in the offer/answer exchange SHOULD be as close to those in the pre-re-INVITE state as possible.
3.5. Glare Situations

Section 4 of RFC 3264 [RFC3264] defines glare conditions as a user agent receiving an offer after having sent one but before having received an answer to it. That section specifies rules to avoid glare situations in most cases. When despite following those rules a glare condition occurs (as a result of a race condition), it is handled as specified in Sections 14.1 and 14.2 of RFC 3261 [RFC3261]. The UAS returns a 491 (Request Pending) response and the UAC retries the offer after a randomly-selected time, which depends on which user agent is the owner of the Call-ID of the dialog. The rules in RFC 3261 [RFC3261] not only cover collisions between re-INVITEs that contain offers; they cover collisions between two re-INVITEs in general, even if they do not contain offers. Sections 5.2 and 5.3 of RFC 3311 [RFC3311] extend those rules to also cover collisions between an UPDATE request carrying an offer and another message (UPDATE, PRACK or INVITE) also carrying an offer.

The rules in RFC 3261 [RFC3261] do not cover collisions between an UPDATE request and a non-2xx final response to a re-INVITE. Since both the UPDATE request and the reliable response could be requesting changes to the session state, it would not be clear which changes would need to be executed first. However, the procedures discussed in Section 3.4 already cover this type of situation. Therefore, there is no need to specify further rules here.

3.6. Example of UAS Behavior

This section contains an example of a UAS that implements this specification using an UPDATE request and a 2xx response to a re-INVITE in order to revert to the pre-re-INVITE state. The example, which is shown in Figure 4, assumes that the UAS requires its user’s input in order to accept or reject the addition of a video stream and uses reliable provisional responses [RFC3262] (PRACK transactions are not shown for clarity).
Figure 4: Rejection of a video stream by the user

The UAs perform an offer/answer exchange to establish an audio only session:

SDP1:
   m=audio 30000 RTP/AVP 0
c=IN IP4 192.0.2.1

SDP2:
   m=audio 31000 RTP/AVP 0
c=IN IP4 192.0.2.5

At a later point, the UAC sends a re-INVITE (4) in order to add a new codec to the audio stream and to add a video stream to the session.
In (5), the UAS accepts the addition of the audio codec but does not accept the video stream yet (it provides a null IP address instead of setting the stream to ‘inactive’ because inactive streams still need to exchange RTCP traffic).

SDP4:
```
m=audio 31000 RTP/AVP 0 3  
c=IN IP4 192.0.2.5  
m=video 31002 RTP/AVP 31  
c=IN IP4 0.0.0.0
```

At a later point, the UAC sends an UPDATE request (6) to remove the original audio codec from the audio stream (the UAC could have also used the PRACK to (5) to request this change).

SDP5:
```
m=audio 30000 RTP/AVP 3  
c=IN IP4 192.0.2.1  
m=video 30002 RTP/AVP 31  
c=IN IP4 192.0.2.1
```

Yet at a later point, the UAS’s user rejects the addition of the video stream. Additionally, the UAS decides to revert to the original audio codec. Consequently, the UAS sends an UPDATE request (8) setting the port of the video stream to zero and offering the original audio codec in its SDP.

SDP7:
```
m=audio 31000 RTP/AVP 0  
c=IN IP4 192.0.2.5  
m=video 0 RTP/AVP 31  
c=IN IP4 0.0.0.0
```
The UAC accepts the change in the audio codec in its 200 (OK) response (9) to the UPDATE request.

SDP8:

m=audio 30000 RTP/AVP 0
c=IN IP4 192.0.2.1
m=video 0 RTP/AVP 31
c=IN IP4 192.0.2.1

The UAS now returns a 200 (OK) response (10) to the re-INVITE. Note that the media state after this 200 (OK) response is the same as the pre-re-INVITE media state.

3.7. Example of UAC Behavior

Figure 5 shows an example of a race condition situation in which the UAs end up with different views of the state of the session.

```
UA1                  Proxy                  UA2
|                      |                      |
|-----(1) INVITE SDP1--|--(2) INVITE SDP1-->    |
a:sendrecv            a:sendrecv          a:sendrecv
v:inactive            v:inactive          v:inactive

<----(4) 183 SDP2-----<----(3) 183 SDP2-----
a:sendrecv            a:sendrecv          a:sendrecv
v:sendonly            v:recvonly          v:inactive

+--(7) 4xx             <-----(5) 4xx ------
|                             <---------(6) ACK ------>
|                             a:sendrecv
|                             v:inactive
<-----(9) UPDATE SDP3--

a:sendonly            a:sendonly          a:recvonly
v:inactive            v:inactive          v:inactive

<-(7) 4xx -            <-(11) 200 OK SDP4-->

-------------(12) ACK -------->

a: status of the audio stream
v: status of the video stream
```
The UAs in Figure 5 are involved in a session that, just before the message flows in the figures starts, includes a sendrecv audio stream and an inactive video stream. UA1 sends a re-INVITE (1) requesting to make the video stream sendrecv.

SDP1:
   m=audio 20000 RTP/AVP 0
   a=sendrecv
   m=video 20002 RTP/AVP 31
   a=sendrecv

UA2 is configured to automatically accept incoming video streams but to ask for user input before generating an outgoing video stream. Therefore, UA2 makes the video stream recvonly by returning a 183 (Session Progress) response (2).

SDP2:
   m=audio 30000 RTP/AVP 0
   a=sendrecv
   m=video 30002 RTP/AVP 31
   a=recvonly

When asked for input, UA2’s user chooses not to have either incoming or outgoing video. In order to make the video stream inactive, UA2 returns a 4xx error response (5) to the re-INVITE. The ACK request (6) for this error response is generated by the proxy between both user agents. Note that this error response undoes already-executed changes. So, UA2 is a legacy UA that does not support this specification.

The proxy relays the 4xx response (7) towards UA1. However, the 4xx response (7) takes time to arrive to UA1 (e.g., the response may have been sent over UDP and the first few retransmissions were lost). In the meantime, UA2’s user decides to put the audio stream on hold. UA2 sends an UPDATE request (8) making the audio stream recvonly. The video stream, which is inactive, is not modified and, thus, continues being inactive.

SDP3:
   m=audio 30000 RTP/AVP 0
   a=recvonly
   m=video 30002 RTP/AVP 31
   a=inactive
The proxy relays the UPDATE request (9) to UA1. The UPDATE request (9) arrives at UA1 before the 4xx response (7) that had been previously sent. UA1 accepts the changes in the UPDATE request and returns a 200 (OK) response (10) to it.

SDP4:
  m=audio 20000 RTP/AVP 0
  a=sendonly
  m=video 30002 RTP/AVP 31
  a=inactive

At a later point, the 4xx response (7) finally arrives at UA1. This response makes the session return to its pre-re-INVITE state. Therefore, for UA1, the audio stream is sendrecv and the video stream is inactive. However, for UA2, the audio stream is recvonly (the video stream is also inactive).

After the message flow in Figure 5, following the recommendations in this section, when UA1 received an error response (7) that undid already-executed changes, UA1 would generate an UPDATE request with an SDP reflecting the pre-re-INVITE state (i.e., sendrecv audio and inactive video). UA2 could then return a 200 (OK) response to the UPDATE request making the audio stream recvonly, which is the state UA2’s user had requested. Such an UPDATE transaction would get the UAs back into synchronization.

3.8. Clarifications on Cancelling Re-INVITEs

Section 9.2 of RFC 3261 [RFC3261] specifies the behavior of a UAS responding to a CANCEL request. Such a UAS responds to the INVITE request with a 487 (Request Terminated) at the ‘should’ level. Per the rules specified in Section 3.3, if the INVITE request was a re-INVITE and some of its requested changes had already been executed, the UAS would return a 2xx response instead.

4. Refreshing a Dialog’s Targets

The following sections discuss how to refresh the targets of a dialog.

4.1. Background and Terminology on a Dialog’s Targets

As described in Section 12 of RFC 3261 [RFC3261], a UA involved in a dialog keeps a record of the SIP or SIPS URI at which it can communicate with a specific instance of its peer (this is called the "dialog’s remote target URI" and is equal to the URI contained in the
This document introduces the complementary concept of the "dialog’s local target URI", defined as a UA’s record of the SIP or SIPS URI at which the peer can communicate with it (equal to the URI contained in the Contact header of requests and responses it sends to the peer). These terms are complementary because the "dialog’s remote target URI" according to one UA is the "dialog’s local target URI" according to the other UA, and vice-versa.

4.2. Background on Target-refresh Requests

A target-refresh request is defined as follows in Section 6 of RFC 3261 [RFC3261]:

"A target-refresh request sent within a dialog is defined as a request that can modify the remote target of the dialog."

Additionally, 2xx responses to target-refresh requests can also update the remote target of the dialog. As discussed in Section 12.2 of RFC 3261 [RFC3261], re-INVITEs are target-refresh requests.

RFC 3261 [RFC3261] specifies the behavior of UASs receiving target-refresh requests and of UACs receiving a 2xx response for a target-refresh request.

Section 12.2.2 of RFC 3261 [RFC3261] says:

"When a UAS receives a target-refresh request, it MUST replace the dialog’s remote target URI with the URI from the Contact header field in that request, if present."

Section 12.2.1.2 of RFC 3261 [RFC3261] says:

"When a UAC receives a 2xx response to a target-refresh request, it MUST replace the dialog’s remote target URI with the URI from the Contact header field in that response, if present."

The fact that re-INVITEs can be long-lived transactions and can have other transactions within them makes it necessary to revise these rules. Section 4.3 specifies new rules for the handling of target-refresh requests. Note that the new rules apply to any target-refresh request, not only to re-INVITEs.

4.3. Clarification on the Atomicity of Target-Refresh Requests

The local and remote targets of a dialog are special types of state information because of their essential role in the exchange of SIP messages between UAs in a dialog. A UA involved in a dialog receives
the remote target of the dialog from the remote UA. The UA uses the received remote target to send SIP requests to the remote UA.

The dialog’s local target is a piece of state information that is not meant to be negotiated. When a UA changes its local target (i.e., the UA changes its IP address), the UA simply communicates its new local target to the remote UA (e.g., the UA communicates its new IP address to the remote UA in order to remain reachable by the remote UA). UAs need to follow the behavior specified in Section 4.4, Section 4.5, Section 4.6, and Section 4.7 of this specification instead of that specified in RFC 3261 [RFC3261], which was discussed in Section 4.2. The new behavior regarding target-refresh requests implies that a target-refresh request can, in some cases, update the remote target even if the request is responded with a final error response. This means that target-refresh requests are not atomic.

4.4. UA Updating the Dialog’s Local Target in a Request

In order to update its local target, a UA can send a target-refresh request. If the UA receives an error response to the target-refresh request, the remote UA has not updated its remote target.

This allows UASs to authenticate target-refresh requests (see Section 26.2 of RFC 3261 [RFC3261]).

If the UA receives a reliable provisional response or a 2xx response to the target-refresh request, or the UA receives an in-dialog request on the new local target, the remote UA has updated its remote target. The UA can consider the target refresh operation completed.

Even if the target request was a re-INVITE and the final response to the re-INVITE was an error response, the UAS would not revert to the pre-re-INVITE remote target.

A UA SHOULD NOT use the same target refresh request to refresh the target and to make session changes unless the session changes can be trivially accepted by the remote UA (e.g., an IP address change). Piggybacking a target refresh with more complicated session changes would make it unnecessarily complicated for the remote UA to accept the target refresh while rejecting the session changes. Only in case the target refresh request is a re-INVITE and the UAS supports reliable provisional response or UPDATE requests, the UAC MAY piggyback session changes and a target refresh in the same re-INVITE.

4.5. UA Updating the Dialog’s Local Target in a Response

A UA processing an incoming target refresh request can update its local target by returning a reliable provisional response or a 2xx
response to the target-refresh request. The response needs to contain the updated local target URI in its Contact header field. On sending the response, the UA can consider the target refresh operation completed.

4.6. A Request Updating the Dialog’s Remote Target

Behavior of a UA after having received a target-refresh request updating the remote target:

If the UA receives a target-refresh request that has been properly authenticated (see Section 26.2 of RFC 3261 [RFC3261]), the UA SHOULD generate a reliable provisional response or a 2xx response to the target-refresh request. If generating such responses is not possible (e.g., the UA does not support reliable provisional responses and needs user input before generating a final response), the UA SHOULD send an in-dialog request to the remote UA using the new remote target (if the UA does not need to send a request for other reasons, the UAS can send an UPDATE request). On sending a reliable provisional response or a 2xx response to the target-refresh request, or a request to the new remote target, the UA MUST replace the dialog’s remote target URI with the URI from the Contact header field in the target-refresh request.

Reliable provisional responses in SIP are specified in RFC 3262 [RFC3262]. In this document, reliable provisional responses are those that use the mechanism defined in RFC 3262 [RFC3262]. Other specifications may define ways to send provisional responses reliably using non-SIP mechanisms (e.g., using media-level messages to acknowledge the reception of the SIP response). For the purposes of this document, provisional responses using those non-SIP mechanisms are considered unreliable responses. Note that non-100 provisional responses are only applicable to INVITE transactions [RFC4320].

If instead of sending a reliable provisional response or a 2xx response to the target-refresh request, or a request to the new target, the UA generates an error response to the target-refresh request, the UA MUST NOT update its dialog’s remote target.

4.7. A Response Updating the Dialog’s Remote Target

If a UA receives a reliable provisional response or a 2xx response to a target-refresh request, the UA MUST replace the dialog’s remote target URI with the URI from the Contact header field in that response, if present.

If a UA receives an unreliable provisional response to a target-
refresh request, the UA MUST NOT refresh the dialog’s remote target.

4.8. Race Conditions and Target Refreshes

SIP provides request ordering by using the Cseq header field. That is, a UA that receives two requests at roughly the same time can know which one is newer. However, SIP does not provide ordering between responses and requests. For example, if a UA receives a 200 (OK) response to an UPDATE request and an UPDATE request at roughly the same time, the UA cannot know which one was sent last. Since both messages can refresh the remote target, the UA needs to know which message was sent last in order to know which remote target needs to be used.

This document specifies the following rule to avoid the situation just described. If the protocol allows a UA to use a target-refresh request at the point in time that UA wishes to refresh its local target, the UA MUST use a target-refresh request instead of a response to refresh its local target. This rule implies that a UA only uses a response (i.e., a reliable provisional response or a 2xx response to a target-refresh request) to refresh its local target if the UA is unable to use a target-refresh request at that point in time (e.g., the UAS of an ongoing re-INVITE without support for UPDATE).

4.9. Early Dialogs

The rules given in this section about which messages can refresh the target of a dialog also apply to early dialogs created by an initial INVITE transaction. Additionally, as specified in Section 13.2.2.4 of RFC 3261 [RFC3261], on receiving a 2xx response to the initial INVITE, the UAC recomputes the whole route set of the dialog, which transitions from the "early" state to the "confirmed" state.

Section 12.1 of RFC 3261 allows unreliable provisional responses to create early dialogs. However, per the rules given in this section, unreliable provisional responses cannot refresh the target of a dialog. Therefore, the UAC of an initial INVITE transaction will not perform any target refresh as a result of the reception of an unreliable provisional response with an updated Contact value on an (already-established) early dialog. Note also that a given UAS can establish additional early dialogs, which can have different targets, by returning additional unreliable provisional responses with different To tags.
5. A UA Losing its Contact

The following sections discuss the case where a UA loses its transport address during an ongoing re-INVITE transaction. Such a UA will refresh the dialog’s local target so that it reflects its new transport address. Note that target refreshes that do not involve changes in the UA’s transport address are outside of the scope of this section. Also, UAs losing their transport address during a non-re-INVITE transaction (e.g., a UA losing its transport address right after having sent an UPDATE request before having received a response to it) are out of scope as well.

The rules given in this section are also applicable to initial INVITE requests that have established early dialogs.

5.1. Background on re-INVITE Transaction Routing

Re-INVITEs are routed using the dialog’s route set, which contains all the proxy servers that need to be traversed by requests sent within the dialog. Responses to the re-INVITE are routed using the Via entries in the re-INVITE.

ACK requests for 2xx responses and for non-2xx final responses are generated in different ways. As specified in Sections 14.1 and 13.2.1 of RFC 3261 [RFC3261], ACK requests for 2xx responses are generated by the UAC core and are routed using the dialog’s route set. As specified in Section 17.1.1.2 of RFC 3261 [RFC3261], ACK requests for non-2xx final responses are generated by the INVITE client transaction (i.e., they are generated in a hop-by-hop fashion by the proxy servers in the path) and are sent to the same transport address as the re-INVITE.

5.2. Problems with UAs Losing their Contact

Refreshing the dialog’s remote target during a re-INVITE transaction (see Section 4.3) presents some issues because of the fact that re-INVITE transactions can be long lived. As described in Section 5.1, the way responses to the re-INVITE and ACKs for non-2xx final responses are routed is fixed once the re-INVITE is sent. The routing of this messages does not depend on the dialog’s route set and, thus, target refreshes within an ongoing re-INVITE do not affect their routing. A UA that changes its location (i.e., performs a target refresh) but is still reachable at its old location will be able to receive those messages (which will be sent to the old location). However, a UA that cannot be reachable at its old location any longer will not be able to receive them.

The following sections describe the errors UAs face when they lose
their transport address during a re-INVITE. On detecting some of these errors, UAs following the rules specified in RFC 3261 [RFC3261] will terminate the dialog. When the dialog is terminated, the only option for the UAs is to establish a new dialog. The following sections change the requirements RFC 3261 [RFC3261] places on UAs when certain errors occur so that the UAs can recover from those errors. In short, the UAs generate a new re-INVITE transaction to synchronize both UAs. Note that there are existing UA implementations deployed that already implement this behavior.

5.3. UAS Losing its Contact: UAC Behavior

When a UAS that moves to a new contact and loses its old contact generates a non-2xx final response to the re-INVITE, it will not be able to receive the ACK request. The entity receiving the response and, thus, generating the ACK request will either get a transport error or a timeout error, which, as described in Section 8.1.3.1 of RFC 3261 [RFC3261], will be treated as a 503 (Service Unavailable) response and as a 408 (Request Timeout) response, respectively. If the sender of the ACK request is a proxy server, it will typically ignore this error. If the sender of the ACK request is the UAC, according to Section 12.2.1.2 of RFC 3261 [RFC3261], it is supposed to (at the "should" level) terminate the dialog by sending a BYE request. However, because of the special properties of ACK requests for non-2xx final responses, most existing UACs do not terminate the dialog when ACK request fails, which is fortunate.

A UAC that accepts a target refresh within a re-INVITE MUST ignore transport and timeout errors when generating an ACK request for a non-2xx final response. Additionally, UAC SHOULD generate a new re-INVITE in order to make sure that both UAs have a common view of the state of the session.

It is possible that the errors ignored by the UAC were not related to the target refresh operation. If that was the case, the second re-INVITE would fail and the UAC would terminate the dialog because, per the rules above, UACs only ignore errors when they accept a target refresh within the re-INVITE.

5.4. UAC Losing its Contact: UAS Behavior

When a UAC moves to a new contact and loses its old contact, it will not be able to receive responses to the re-INVITE. Consequently, it will never generate an ACK request.

As described in Section 16.9 of RFC 3261 [RFC3261], a proxy server that gets an error when forwarding a response does not take any measures. Consequently, proxy servers relaying responses will...
effectively ignore the error.

If there are no proxy servers in the dialog’s route set, the UAS will get an error when sending a non-2xx final response. The UAS core will be notified of the transaction failure, as described in Section 17.2.1 of RFC 3261 [RFC3261]. Most existing UASs do not terminate the dialog on encountering this failure, which is fortunate.

Regardless of the presence or absence of proxy servers in the dialog’s route set, a UAS generating a 2xx response to the re-INVITE will never receive an ACK request for it. According to Section 14.2 of RFC 3261 [RFC3261], such a UAS is supposed to (at the "should" level) terminate the dialog by sending a BYE request.

A UAS that accepts a target refresh within a re-INVITE and never receives an ACK request after having sent a final response to the re-INVITE SHOULD NOT terminate the dialog if the UA has received a new re-INVITE with a higher CSeq sequence number than the original one.

5.5. UAC Losing its Contact: UAC Behavior

When a UAC moves to a new contact and loses its old contact, it will not be able to receive responses to the re-INVITE. Consequently, it will never generate an ACK request.

Such a UAC SHOULD generate a CANCEL request to cancel the re-INVITE and cause the INVITE client transaction corresponding to the re-INVITE to enter the "Terminated" state. The UAC SHOULD also send a new re-INVITE in order to make sure that both UAs have a common view of the state of the session.

Per Section 14.2 of RFC 3261 [RFC3261], the UAS will accept new incoming re-INVITES as soon as it has generated a final response to the previous INVITE request, which had a lower CSeq sequence number.

6. Security Considerations

This document does not introduce any new security issue. It just clarifies how certain transactions should be handled in SIP. Security issues related to re-INVITEs and UPDATE requests are discussed in RFC 3261 [RFC3261] and RFC 3311 [RFC3311].

In particular, in order not to reduce the security level for a given session, re-INVITEs and UPDATE requests SHOULD be secured using a mechanism equivalent to or stronger than the initial INVITE request that created the session. For example, if the initial INVITE request
was end-to-end integrity protected or encrypted, subsequent re-
INVITEs and UPDATE requests should also be so.

7. IANA Considerations

There are no IANA actions associated with this document.

8. Acknowledgements

Paul Kyzivat provided useful ideas on the topics discussed in this
document.

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate

[RFC3261] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston,
A., Peterson, J., Sparks, R., Handley, M., and E.
Schooler, "SIP: Session Initiation Protocol", RFC 3261,
June 2002.

[RFC3262] Rosenberg, J. and H. Schulzrinne, "Reliability of
Provisional Responses in Session Initiation Protocol
(SIP)", RFC 3262, June 2002.

with Session Description Protocol (SDP)", RFC 3264,
June 2002.


[RFC4032] Camarillo, G. and P. Kyzivat, "Update to the Session
Initiation Protocol (SIP) Preconditions Framework",

9.2. Informative References

[RFC4320] Sparks, R., "Actions Addressing Identified Issues with the
Session Initiation Protocol’s (SIP) Non-INVITE

Authors’ Addresses

Gonzalo Camarillo (editor)
Ericsson
Hirsalantie 11
Jorvas 02420
Finland

Email: Gonzalo.Camarillo@ericsson.com

Christer Holmberg
Ericsson
Hirsalantie 11
Jorvas 02420
Finland

Email: Christer.Holmberg@ericsson.com

Yang Gao
ZTE
China

Email: gao.yang2@zte.com.cn
This document describes an extension to the Session Initiation Protocol (SIP) defined by RFC 3261. The purpose of this extension is to provide an extensible framework by which SIP nodes can request notification from remote nodes indicating that certain events have occurred.

Note that the event notification mechanisms defined herein are NOT intended to be a general-purpose infrastructure for all classes of event subscription and notification.

This document represents a backwards-compatible improvement on the original mechanism described by RFC 3265, taking into account several years of implementation experience. Accordingly, this document obsoletes RFC 3265. This document also updates RFC 4660 slightly to accommodate some small changes to the mechanism that were discussed in that document.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on November 1, 2012.
Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the
document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal
Provisions Relating to IETF Documents
(http://trustee.ietf.org/license-info) in effect on the date of
publication of this document. Please review these documents
carefully, as they describe your rights and restrictions with respect
to this document. Code Components extracted from this document must
include Simplified BSD License text as described in Section 4.e of
the Trust Legal Provisions and are provided without warranty as
described in the Simplified BSD License.

Table of Contents

1. Introduction .......................... 5
   1.1. Overview of Operation ............... 5
   1.2. Documentation Conventions .......... 6
2. Definitions .......................... 6
3. SIP Methods for Event Notification .... 7
   3.1. SUBSCRIBE .......................... 7
   3.1.1. Subscription Duration ............. 7
   3.1.2. Identification of Subscribed Events and Event
          Classes .......................... 8
   3.1.3. Additional SUBSCRIBE Header Field Values ..... 9
   3.2. NOTIFY ............................ 9
   3.2.1. Identification of Reported Events, Event Classes,
          and Current State .................. 9
4. Node Behavior ....................... 10
   4.1. Subscriber Behavior ............... 10
   4.1.1. Detecting Support for SIP Events .... 10
   4.1.2. Creating and Maintaining Subscriptions .... 10
   4.1.3. Receiving and Processing State Information ..... 14
   4.1.4. Forking of SUBSCRIBE Requests ..... 16
   4.2. Notifier Behavior ................ 17
   4.2.1. Subscription Establishment and Maintenance .... 17
   4.2.2. Sending State Information to Subscribers .... 20
   4.2.3. PINT Compatibility ................ 23
   4.3. Proxy Behavior .................. 23
   4.4. Common Behavior ................. 23
   4.4.1. Dialog Creation and Termination .... 24
   4.4.2. Notifier Migration ............. 24
   4.4.3. Polling Resource State ........... 25
   4.4.4. Allow-Events header field usage .... 26
4.5. Targeting Subscriptions at Devices .. 26
Appendix B. Changes from RFC 3265

B.1. Bug 666: Clarify use of expires=xxx with terminated

B.2. Bug 667: Reason code for unsub/poll not clearly
spelled out

B.3. Bug 669: Clarify: SUBSCRIBE for a duration might be
answered with a NOTIFY/expires=0

B.4. Bug 670: Dialog State Machine needs clarification

B.5. Bug 671: Clarify timeout-based removal of subscriptions

B.6. Bug 672: Mandate expires= in NOTIFY

B.7. Bug 673: INVITE 481 response effect clarification

B.8. Bug 677: SUBSCRIBE response matching text in error

B.9. Bug 695: Document is not explicit about response to
NOTIFY at subscription termination

B.10. Bug 696: Subscription state machine needs clarification

B.11. Bug 697: Unsubscription behavior could be clarified

B.12. Bug 699: NOTIFY and SUBSCRIBE are target refresh
requests

B.13. Bug 722: Inconsistent 423 reason phrase text

B.14. Bug 741: guidance needed on when to not include
Allow-Events

B.15. Bug 744: 5xx to NOTIFY terminates a subscription, but
should not

B.16. Bug 752: Detection of forked requests is incorrect

B.17. Bug 773: Reason code needs IANA registry

B.18. Bug 774: Need new reason for terminating subscriptions
to resources that never change

B.19. Clarify handling of Route/Record-Route in NOTIFY

B.20. Eliminate implicit subscriptions

B.21. Deprecate dialog re-use

B.22. Rationalize dialog creation

B.23. Refactor behavior sections

B.24. Clarify sections that need to be present in event
packages

B.25. Make CANCEL handling more explicit

B.26. Remove State Agent Terminology

B.27. Miscellaneous Changes

Appendix A. Acknowledgements

B.1. Bug 666: Clarify use of expires=xxx with terminated

B.2. Bug 667: Reason code for unsub/poll not clearly
spelled out

B.3. Bug 669: Clarify: SUBSCRIBE for a duration might be
answered with a NOTIFY/expires=0

B.4. Bug 670: Dialog State Machine needs clarification

B.5. Bug 671: Clarify timeout-based removal of subscriptions

B.6. Bug 672: Mandate expires= in NOTIFY

B.7. Bug 673: INVITE 481 response effect clarification

B.8. Bug 677: SUBSCRIBE response matching text in error

B.9. Bug 695: Document is not explicit about response to
NOTIFY at subscription termination

B.10. Bug 696: Subscription state machine needs clarification

B.11. Bug 697: Unsubscription behavior could be clarified

B.12. Bug 699: NOTIFY and SUBSCRIBE are target refresh
requests

B.13. Bug 722: Inconsistent 423 reason phrase text

B.14. Bug 741: guidance needed on when to not include
Allow-Events

B.15. Bug 744: 5xx to NOTIFY terminates a subscription, but
should not

B.16. Bug 752: Detection of forked requests is incorrect

B.17. Bug 773: Reason code needs IANA registry

B.18. Bug 774: Need new reason for terminating subscriptions
to resources that never change

B.19. Clarify handling of Route/Record-Route in NOTIFY

B.20. Eliminate implicit subscriptions

B.21. Deprecate dialog re-use

B.22. Rationalize dialog creation

B.23. Refactor behavior sections

B.24. Clarify sections that need to be present in event
packages

B.25. Make CANCEL handling more explicit

B.26. Remove State Agent Terminology

B.27. Miscellaneous Changes

Author’s Address
1. Introduction

The ability to request asynchronous notification of events proves useful in many types of SIP services for which cooperation between end-nodes is required. Examples of such services include automatic callback services (based on terminal state events), buddy lists (based on user presence events), message waiting indications (based on mailbox state change events), and PSTN and Internet Internetworking (PINT) [RFC2848] status (based on call state events).

The methods described in this document provide a framework by which notification of these events can be ordered.

The event notification mechanisms defined herein are NOT intended to be a general-purpose infrastructure for all classes of event subscription and notification. Meeting requirements for the general problem set of subscription and notification is far too complex for a single protocol. Our goal is to provide a SIP-specific framework for event notification which is not so complex as to be unusable for simple features, but which is still flexible enough to provide powerful services. Note, however, that event packages based on this framework may define arbitrarily elaborate rules which govern the subscription and notification for the events or classes of events they describe.

This document does not describe an extension which may be used directly; it must be extended by other documents (herein referred to as "event packages"). In object-oriented design terminology, it may be thought of as an abstract base class which must be derived into an instantiatable class by further extensions. Guidelines for creating these extensions are described in Section 5.

1.1. Overview of Operation

The general concept is that entities in the network can subscribe to resource or call state for various resources or calls in the network, and those entities (or entities acting on their behalf) can send notifications when those states change.

A typical flow of messages would be:

<table>
<thead>
<tr>
<th>Subscriber</th>
<th>Notifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>-----SUBSCRIBE-----</td>
<td>Request state subscription</td>
</tr>
<tr>
<td>&lt;-----200---------</td>
<td>Acknowledge subscription</td>
</tr>
<tr>
<td>&lt;-----NOTIFY------</td>
<td>Return current state information</td>
</tr>
<tr>
<td>---200---------</td>
<td>Return current state information</td>
</tr>
<tr>
<td>&lt;-----NOTIFY------</td>
<td>Return current state information</td>
</tr>
<tr>
<td>---200---------</td>
<td></td>
</tr>
</tbody>
</table>
Subscriptions are expired and must be refreshed by subsequent SUBSCRIBE requests.

1.2. Documentation Conventions

There are several paragraphs throughout this document which provide motivational or clarifying text. Such passages are non-normative, and are provided only to assist with reader comprehension. These passages are set off from the remainder of the text by being indented thus:

This is an example of non-normative explanatory text. It does not form part of the specification, and is used only for clarification.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

In particular, implementors need to take careful note of the meaning of "SHOULD" defined in RFC 2119. To rephrase: violation of SHOULD-strength requirements requires careful analysis and clearly enumerable reasons. It is a protocol violation to fail to comply with "SHOULD"-strength requirements whimsically or for ease of implementation.

The use of quotation marks next to periods and commas follows the convention used by the American Mathematical Society; although contrary to traditional American English convention, this usage lends clarity to certain passages.

2. Definitions

Event Package: An event package is an additional specification which defines a set of state information to be reported by a notifier to a subscriber. Event packages also define further syntax and semantics based on the framework defined by this document required to convey such state information.

Event Template-Package: An event template-package is a special kind of event package which defines a set of states which may be applied to all possible event packages, including itself.
Notification: Notification is the act of a notifier sending a NOTIFY request to a subscriber to inform the subscriber of the state of a resource.

Notifier: A notifier is a user agent which generates NOTIFY requests for the purpose of notifying subscribers of the state of a resource. Notifiers typically also accept SUBSCRIBE requests to create subscriptions.

Subscriber: A subscriber is a user agent which receives NOTIFY requests from notifiers; these NOTIFY requests contain information about the state of a resource in which the subscriber is interested. Subscribers typically also generate SUBSCRIBE requests and send them to notifiers to create subscriptions.

Subscription: A subscription is a set of application state associated with a dialog. This application state includes a pointer to the associated dialog, the event package name, and possibly an identification token. Event packages will define additional subscription state information. By definition, subscriptions exist in both a subscriber and a notifier.

Subscription Migration: Subscription migration is the act of moving a subscription from one notifier to another notifier.

3. SIP Methods for Event Notification

3.1. SUBSCRIBE

The SUBSCRIBE method is used to request current state and state updates from a remote node. SUBSCRIBE requests are target refresh requests, as that term is defined in [RFC3261].

3.1.1. Subscription Duration

SUBSCRIBE requests SHOULD contain an "Expires" header field (defined in [RFC3261]). This expires value indicates the duration of the subscription. In order to keep subscriptions effective beyond the duration communicated in the "Expires" header field, subscribers need to refresh subscriptions on a periodic basis using a new SUBSCRIBE request on the same dialog as defined in [RFC3261].

If no "Expires" header field is present in a SUBSCRIBE request, the implied default MUST be defined by the event package being used.

200-class responses to SUBSCRIBE requests also MUST contain an "Expires" header field. The period of time in the response MAY be
shorter but MUST NOT be longer than specified in the request. The notifier is explicitly allowed to shorten the duration to zero. The period of time in the response is the one which defines the duration of the subscription.

An "expires" parameter on the "Contact" header field has no semantics for the SUBSCRIBE method and is explicitly not equivalent to an "Expires" header field in a SUBSCRIBE request or response.

A natural consequence of this scheme is that a SUBSCRIBE request with an "Expires" of 0 constitutes a request to unsubscribe from the matching subscription.

In addition to being a request to unsubscribe, a SUBSCRIBE request with "Expires" of 0 also causes a fetch of state; see Section 4.4.3.

Notifiers may also wish to cancel subscriptions to events; this is useful, for example, when the resource to which a subscription refers is no longer available. Further details on this mechanism are discussed in Section 4.2.2.

3.1.2. Identification of Subscribed Events and Event Classes

Identification of events is provided by three pieces of information: Request URI, Event Type, and (optionally) message body.

The Request URI of a SUBSCRIBE request, most importantly, contains enough information to route the request to the appropriate entity per the request routing procedures outlined in [RFC3261]. It also contains enough information to identify the resource for which event notification is desired, but not necessarily enough information to uniquely identify the nature of the event (e.g., "sip:adam@example.com" would be an appropriate URI to subscribe to for my presence state; it would also be an appropriate URI to subscribe to the state of my voice mailbox).

Subscribers MUST include exactly one "Event" header field in SUBSCRIBE requests, indicating to which event or class of events they are subscribing. The "Event" header field will contain a token which indicates the type of state for which a subscription is being requested. This token will be registered with the IANA and will correspond to an event package which further describes the semantics of the event or event class.

If the event package to which the event token corresponds defines behavior associated with the body of its SUBSCRIBE requests, those semantics apply.
Event packages may also define parameters for the Event header field; if they do so, they must define the semantics for such parameters.

3.1.3. Additional SUBSCRIBE Header Field Values

Because SUBSCRIBE requests create a dialog usage as defined in [RFC3261], they MAY contain an "Accept" header field. This header field, if present, indicates the body formats allowed in subsequent NOTIFY requests. Event packages MUST define the behavior for SUBSCRIBE requests without "Accept" header fields; usually, this will connote a single, default body type.

Header values not described in this document are to be interpreted as described in [RFC3261].

3.2. NOTIFY

NOTIFY requests are sent to inform subscribers of changes in state to which the subscriber has a subscription. Subscriptions are created using the SUBSCRIBE method. In legacy implementations, it is possible that other means of subscription creation have been used. However, this specification does not allow the creation of subscriptions except through SUBSCRIBE requests and (for backwards-compatibility) REFER requests [RFC3515].

NOTIFY is a target refresh request, as that term is defined in [RFC3261].

A NOTIFY request does not terminate its corresponding subscription; in other words, a single SUBSCRIBE request may trigger several NOTIFY requests.

3.2.1. Identification of Reported Events, Event Classes, and Current State

Identification of events being reported in a notification is very similar to that described for subscription to events (see Section 3.1.2).

As in SUBSCRIBE requests, NOTIFY request "Event" header fields MUST contain a single event package name for which a notification is being generated. The package name in the "Event" header field MUST match the "Event" header field in the corresponding SUBSCRIBE request.

Event packages may define semantics associated with the body of their NOTIFY requests; if they do so, those semantics apply. NOTIFY request bodies are expected to provide additional details about the nature of the event which has occurred and the resultant resource
When present, the body of the NOTIFY request MUST be formatted into one of the body formats specified in the "Accept" header field of the corresponding SUBSCRIBE request (or the default type according to the event package description, if no Accept header field was specified). This body will contain either the state of the subscribed resource or a pointer to such state in the form of a URI (see Section 5.4.13).

4. Node Behavior

4.1. Subscriber Behavior

4.1.1. Detecting Support for SIP Events

The extension described in this document does not make use of the "Require" or "Proxy-Require" header fields; similarly, there is no token defined for "Supported" header fields. Potential subscribers may probe for the support of SIP Events using the OPTIONS request defined in [RFC3261].

The presence of "SUBSCRIBE" in the "Allow" header field of any request or response indicates support for SIP Events; further, in the absence of an "Allow" header field, the simple presence of an "Allow-Events" header field is sufficient to indicate that the node that sent the message is capable of acting as a notifier (see Section 4.4.4).

The "methods" parameter for Contact may also be used to specifically announce support for SUBSCRIBE and NOTIFY requests when registering. (See [RFC3840] for details on the "methods" parameter).

4.1.2. Creating and Maintaining Subscriptions

From the subscriber’s perspective, a subscription proceeds according to the following state diagram. Events which result in a transition back to the same state are not represented in this diagram.
In the state diagram, "Re-subscription times out" means that an attempt to refresh or update the subscription using a new SUBSCRIBE request does not result in a NOTIFY request before the corresponding Timer N expires.

Any transition from "notify_wait" into a "pending" or "active" state results in a new subscription. Note that multiple subscriptions can be generated as the result of a single SUBSCRIBE request (see Section 4.4.1). Each of these new subscriptions exists in its own independent state machine, and runs its own set of timers.
4.1.2.1. Requesting a Subscription

SUBSCRIBE is a dialog-creating method, as described in [RFC3261].

When a subscriber wishes to subscribe to a particular state for a resource, it forms a SUBSCRIBE request. If the initial SUBSCRIBE request represents a request outside of a dialog (as it typically will), its construction follows the procedures outlined in [RFC3261] for UAC request generation outside of a dialog.

This SUBSCRIBE request will be confirmed with a final response. 200-class responses indicate that the subscription has been accepted, and that a NOTIFY request will be sent immediately.

The "Expires" header field in a 200-class response to SUBSCRIBE request indicates the actual duration for which the subscription will remain active (unless refreshed). The received value might be smaller than the value indicated in the SUBSCRIBE request, but cannot be larger; see Section 4.2.1 for details.

Non-200 class final responses indicate that no subscription or new dialog usage has been created, and no subsequent NOTIFY request will be sent. All non-200 class responses (with the exception of "489", described herein) have the same meanings and handling as described in [RFC3261]. For the sake of clarity: if a SUBSCRIBE request contains an "Accept" header field, but that field does not indicate a media type that the notifier is capable of generating in its NOTIFY requests, then the proper error response is 406 (Not Acceptable).

4.1.2.2. Refreshing of Subscriptions

At any time before a subscription expires, the subscriber may refresh the timer on such a subscription by sending another SUBSCRIBE request on the same dialog as the existing subscription. The handling for such a request is the same as for the initial creation of a subscription except as described below.

If a SUBSCRIBE request to refresh a subscription receives a 404, 405, 410, 416, 480-485, 489, 501, or 604 response, the subscriber MUST consider the subscription terminated. (See [RFC5057] for further details and notes about the effect of error codes on dialogs and usages within dialog, such as subscriptions). If the subscriber wishes to re-subscribe to the state, he does so by composing an unrelated initial SUBSCRIBE request with a freshly-generated Call-ID and a new, unique "From" tag (see Section 4.1.2.1.)

If a SUBSCRIBE request to refresh a subscription fails with any error code other than those listed above, the original subscription is
still considered valid for the duration of the most recently known "Expires" value as negotiated by the most recent successful SUBSCRIBE transaction, or as communicated by a NOTIFY request in its "Subscription-State" header field "expires" parameter.

Note that many such errors indicate that there may be a problem with the network or the notifier such that no further NOTIFY requests will be received.

When refreshing a subscription, a subscriber starts Timer N, set to 64*T1, when it sends the SUBSCRIBE request. If this Timer N expires prior to the receipt of a NOTIFY request, the subscriber considers the subscription terminated. If the subscriber receives a success response to the SUBSCRIBE request that indicates that no NOTIFY request will be generated -- such as the 204 response defined for use with the optional extension described in [RFC5839] -- then it MUST cancel Timer N.

4.1.2.3. Unsubscribing

Unsubscribing is handled in the same way as refreshing of a subscription, with the "Expires" header field set to "0". Note that a successful unsubscription will also trigger a final NOTIFY request.

The final NOTIFY request may or may not contain information about the state of the resource; subscribers need to be prepared to receive final NOTIFY requests both with and without state.

4.1.2.4. Confirmation of Subscription Creation

The subscriber can expect to receive a NOTIFY request from each node which has processed a successful subscription or subscription refresh. To ensure that subscribers do not wait indefinitely for a subscription to be established, a subscriber starts a Timer N, set to 64*T1, when it sends a SUBSCRIBE request. If this Timer N expires prior to the receipt of a NOTIFY request, the subscriber considers the subscription failed, and cleans up any state associated with the subscription attempt.

Until Timer N expires, several NOTIFY requests may arrive from different destinations (see Section 4.4.1). Each of these requests establish a new dialog usage and a new subscription. After the expiration of Timer N, the subscriber SHOULD reject any such NOTIFY requests that would otherwise establish a new dialog usage with a "481" response code.

Until the first NOTIFY request arrives, the subscriber should consider the state of the subscribed resource to be in a neutral
state. Event package specifications MUST define this "neutral state" in such a way that makes sense for their application (see Section 5.4.7).

Due to the potential for out-of-order messages, packet loss, and forking, the subscriber MUST be prepared to receive NOTIFY requests before the SUBSCRIBE transaction has completed.

Except as noted above, processing of this NOTIFY request is the same as in Section 4.1.3.

4.1.3. Receiving and Processing State Information

Subscribers receive information about the state of a resource to which they have subscribed in the form of NOTIFY requests.

Upon receiving a NOTIFY request, the subscriber should check that it matches at least one of its outstanding subscriptions; if not, it MUST return a "481 Subscription does not exist" response unless another 400- or 500-class response is more appropriate. The rules for matching NOTIFY requests with subscriptions that create a new dialog usage are described in Section 4.4.1. Notifications for subscriptions which were created inside an existing dialog match if they are in the same dialog and the "Event" header fields match (as described in Section 8.2.1).

If, for some reason, the event package designated in the "Event" header field of the NOTIFY request is not supported, the subscriber will respond with a "489 Bad Event" response.

To prevent spoofing of events, NOTIFY requests SHOULD be authenticated, using any defined SIP authentication mechanism, such as those described in sections 22.2 and 23 of [RFC3261].

NOTIFY requests MUST contain "Subscription-State" header fields which indicate the status of the subscription.

If the "Subscription-State" header field value is "active", it means that the subscription has been accepted and (in general) has been authorized. If the header field also contains an "expires" parameter, the subscriber SHOULD take it as the authoritative subscription duration and adjust accordingly. The "retry-after" and "reason" parameters have no semantics for "active".

If the "Subscription-State" value is "pending", the subscription has been received by the notifier, but there is insufficient policy information to grant or deny the subscription yet. If the header field also contains an "expires" parameter, the subscriber SHOULD
take it as the authoritative subscription duration and adjust accordingly. No further action is necessary on the part of the subscriber. The "retry-after" and "reason" parameters have no semantics for "pending".

If the "Subscription-State" value is "terminated", the subscriber MUST consider the subscription terminated. The "expires" parameter has no semantics for "terminated" -- notifiers SHOULD NOT include an "expires" parameter on a "Subscription-State" header field with a value of "terminated," and subscribers MUST ignore any such parameter, if present. If a reason code is present, the client should behave as described below. If no reason code or an unknown reason code is present, the client MAY attempt to re-subscribe at any time (unless a "retry-after" parameter is present, in which case the client SHOULD NOT attempt re-subscription until after the number of seconds specified by the "retry-after" parameter). The reason codes defined by this document are:

deactivated: The subscription has been terminated, but the subscriber SHOULD retry immediately with a new subscription. One primary use of such a status code is to allow migration of subscriptions between nodes. The "retry-after" parameter has no semantics for "deactivated".

probation: The subscription has been terminated, but the client SHOULD retry at some later time (as long as the resource’s state is still relevant to the client at that time). If a "retry-after" parameter is also present, the client SHOULD wait at least the number of seconds specified by that parameter before attempting to re-subscribe.

rejected: The subscription has been terminated due to change in authorization policy. Clients SHOULD NOT attempt to re-subscribe. The "retry-after" parameter has no semantics for "rejected".

timeout: The subscription has been terminated because it was not refreshed before it expired. Clients MAY re-subscribe immediately. The "retry-after" parameter has no semantics for "timeout". This reason code is also associated with polling of resource state, as detailed in Section 4.4.3

giveup: The subscription has been terminated because the notifier could not obtain authorization in a timely fashion. If a "retry-after" parameter is also present, the client SHOULD wait at least the number of seconds specified by that parameter before attempting to re-subscribe; otherwise, the client MAY retry immediately, but will likely get put back into pending state.
noresource: The subscription has been terminated because the resource state which was being monitored no longer exists. Clients SHOULD NOT attempt to re-subscribe. The "retry-after" parameter has no semantics for "noresource".

invariant: The subscription has been terminated because the resource state is guaranteed not to change for the foreseeable future. This may be the case, for example, when subscribing to the location information of a fixed-location land-line telephone. When using this reason code, notifiers are advised to include a "retry-after" parameter with a large value (for example, 31536000 -- or one year) to prevent older, RFC 3265-compliant clients from periodically resubscribing. Clients SHOULD NOT attempt to resubscribe after receiving a reason code of "invariant," regardless of the presence of or value of a "retry-after" parameter.

Other specifications may define new reason codes for use with the "Subscription-State" header field.

Once the notification is deemed acceptable to the subscriber, the subscriber SHOULD return a 200 response. In general, it is not expected that NOTIFY responses will contain bodies; however, they MAY, if the NOTIFY request contained an "Accept" header field.

Other responses defined in [RFC3261] may also be returned, as appropriate. In no case should a NOTIFY transaction extend for any longer than the time necessary for automated processing. In particular, subscribers MUST NOT wait for a user response before returning a final response to a NOTIFY request.

4.1.4. Forking of SUBSCRIBE Requests

In accordance with the rules for proxying non-INVITE requests as defined in [RFC3261], successful SUBSCRIBE requests will receive only one 200-class response; however, due to forking, the subscription may have been accepted by multiple nodes. The subscriber MUST therefore be prepared to receive NOTIFY requests with "From:" tags which differ from the "To:" tag received in the SUBSCRIBE 200-class response.

If multiple NOTIFY requests are received in different dialogs in response to a single SUBSCRIBE request, each dialog represents a different destination to which the SUBSCRIBE request was forked. Subscriber handling in such situations varies by event package; see Section 5.4.9 for details.
4.2. Notifier Behavior

4.2.1. Subscription Establishment and Maintenance

Notifiers learn about subscription requests by receiving SUBSCRIBE requests from interested parties. Notifiers MUST NOT create subscriptions except upon receipt of a SUBSCRIBE request. However, for historical reasons, the implicit creation of subscriptions as defined in [RFC3515] is still permitted.

[RFC3265] allowed the creation of subscriptions using means other than the SUBSCRIBE method. The only standardized use of this mechanism is the REFER method [RFC3515]. Implementation experience with REFER has shown that the implicit creation of a subscription has a number of undesirable effects, such as the inability to signal the success of a REFER request while signaling a problem with the subscription; and difficulty performing one action without the other. Additionally, the proper exchange of dialog identifiers is difficult without dialog re-use (which has its own set of problems; see Section 4.5).

4.2.1.1. Initial SUBSCRIBE Transaction Processing

In no case should a SUBSCRIBE transaction extend for any longer than the time necessary for automated processing. In particular, notifiers MUST NOT wait for a user response before returning a final response to a SUBSCRIBE request.

This requirement is imposed primarily to prevent the non-INVITE transaction timeout timer F (see [RFC3261]) from firing during the SUBSCRIBE transaction, since interaction with a user would often exceed 64*T1 seconds.

The notifier SHOULD check that the event package specified in the "Event" header field is understood. If not, the notifier SHOULD return a "489 Bad Event" response to indicate that the specified event/event class is not understood.

The notifier SHOULD also perform any necessary authentication and authorization per its local policy. See Section 4.2.1.3.

The notifier MAY also check that the duration in the "Expires" header field is not too small. If and only if the expiration interval is greater than zero AND smaller than one hour AND less than a notifier-configured minimum, the notifier MAY return a "423 Interval Too Brief" error which contains a "Min-Expires" header field field. The "Min-Expires" header field is described in [RFC3261].
Once the notifier determines that it has enough information to create the subscription (i.e., it understands the event package, the subscription pertains to a known resource, and there are no other barriers to creating the subscription), it creates the subscription and a dialog usage, and returns a 200 (OK) response.

When a subscription is created in the notifier, it stores the event package name as part of the subscription information.

The "Expires" values present in SUBSCRIBE 200-class responses behave in the same way as they do in REGISTER responses: the server MAY shorten the interval, but MUST NOT lengthen it.

If the duration specified in a SUBSCRIBE request is unacceptably short, the notifier may be able to send a 423 response, as described earlier in this section.

200-class responses to SUBSCRIBE requests will not generally contain any useful information beyond subscription duration; their primary purpose is to serve as a reliability mechanism. State information will be communicated via a subsequent NOTIFY request from the notifier.

The other response codes defined in [RFC3261] may be used in response to SUBSCRIBE requests, as appropriate.

4.2.1.2. Confirmation of Subscription Creation/Refreshing

Upon successfully accepting or refreshing a subscription, notifiers MUST send a NOTIFY request immediately to communicate the current resource state to the subscriber. This NOTIFY request is sent on the same dialog as created by the SUBSCRIBE response. If the resource has no meaningful state at the time that the SUBSCRIBE request is processed, this NOTIFY request MAY contain an empty or neutral body. See Section 4.2.2 for further details on NOTIFY request generation.

Note that a NOTIFY request is always sent immediately after any 200-class response to a SUBSCRIBE request, regardless of whether the subscription has already been authorized.

4.2.1.3. Authentication/Authorization of SUBSCRIBE Requests

Privacy concerns may require that notifiers apply policy to determine whether a particular subscriber is authorized to subscribe to a certain set of events. Such policy may be defined by mechanisms such as access control lists or real-time interaction with a user. In general, authorization of subscribers prior to authentication is not particularly useful.
SIP authentication mechanisms are discussed in [RFC3261]. Note that, even if the notifier node typically acts as a proxy, authentication for SUBSCRIBE requests will always be performed via a "401" response, not a "407". Notifiers always act as a user agents when accepting subscriptions and sending notifications.

Of course, when acting as a proxy, a node will perform normal proxy authentication (using 407). The foregoing explanation is a reminder that notifiers are always UAs, and as such perform UA authentication.

If authorization fails based on an access list or some other automated mechanism (i.e., it can be automatically authoritatively determined that the subscriber is not authorized to subscribe), the notifier SHOULD reply to the request with a "403 Forbidden" or "603 Decline" response, unless doing so might reveal information that should stay private; see Section 6.2.

If the notifier owner is interactively queried to determine whether a subscription is allowed, a 200 (OK) response is returned immediately. Note that a NOTIFY request is still formed and sent under these circumstances, as described in the previous section.

If subscription authorization was delayed and the notifier wishes to convey that such authorization has been declined, it may do so by sending a NOTIFY request containing a "Subscription-State" header field with a value of "terminated" and a reason parameter of "rejected".

4.2.1.4. Refreshing of Subscriptions

When a notifier receives a subscription refresh, assuming that the subscriber is still authorized, the notifier updates the expiration time for subscription. As with the initial subscription, the server MAY shorten the amount of time until expiration, but MUST NOT increase it. The final expiration time is placed in the "Expires" header field in the response. If the duration specified in a SUBSCRIBE request is unacceptably short, the notifier SHOULD respond with a "423 Interval Too Brief" response.

If no refresh for a notification address is received before its expiration time, the subscription is removed. When removing a subscription, the notifier SHOULD send a NOTIFY request with a "Subscription-State" value of "terminated" to inform it that the subscription is being removed. If such a request is sent, the "Subscription-State" header field SHOULD contain a "reason=timeout" parameter.
Clients can cause a subscription to be terminated immediately by sending a SUBSCRIBE request with an "Expires" header field set to '0'. Notifiers largely treat this the same way as any other subscription expiration: they send a NOTIFY request containing a "Subscription-State" of "terminated", with a reason code of "timeout." For consistency with state polling (see Section 4.4.3) and subscription refreshes, the notifier may choose to include resource state in this final NOTIFY request. However, in some cases, including such state makes no sense. Under such circumstances, the notifier may choose to omit state information from the terminal NOTIFY request.

The sending of a NOTIFY request when a subscription expires allows the corresponding dialog usage to be terminated, if appropriate.

4.2.2. Sending State Information to Subscribers

Notifiers use the NOTIFY method to send information about the state of a resource to subscribers. The notifier’s view of a subscription is shown in the following state diagram. Events which result in a transition back to the same state are not represented in this diagram.
When a SUBSCRIBE request is answered with a 200-class response, the notifier MUST immediately construct and send a NOTIFY request to the subscriber. When a change in the subscribed state occurs, the notifier SHOULD immediately construct and send a NOTIFY request, unless the state transition is caused by a NOTIFY transaction failure. The sending of this NOTIFY message is also subject to authorization, local policy, and throttling considerations.

If the NOTIFY request fails due to expiration of SIP Timer F (transaction timeout), the notifier SHOULD remove the subscription.

This behavior prevents unnecessary transmission of state information for subscribers who have crashed or disappeared from the network. Because such transmissions will be sent multiple times, per the retransmission algorithm defined in [RFC3261] (instead of the typical single transmission for functioning clients), continuing to service them when no client is available...
to acknowledge them could place undue strain on a network. Upon client restart or reestablishment of a network connection, it is expected that clients will send SUBSCRIBE requests to refresh potentially stale state information; such requests will re-install subscriptions in all relevant nodes.

If the NOTIFY transaction fails due to the receipt of a 404, 405, 410, 416, 480-485, 489, 501, or 604 response to the NOTIFY request, the notifier MUST remove the corresponding subscription. See [RFC5057] for further details and notes about the effect of error codes on dialogs and usages within dialog (such as subscriptions).

A notify error response would generally indicate that something has gone wrong with the subscriber or with some proxy on the way to the subscriber. If the subscriber is in error, it makes the most sense to allow the subscriber to rectify the situation (by re-subscribing) once the error condition has been handled. If a proxy is in error, the periodic sending of SUBSCRIBE requests to refresh the expiration timer will re-install subscription state once the network problem has been resolved.

NOTIFY requests MUST contain a "Subscription-State" header field with a value of "active", "pending", or "terminated". The "active" value indicates that the subscription has been accepted and has been authorized (in most cases; see Section 6.2). The "pending" value indicates that the subscription has been received, but that policy information is insufficient to accept or deny the subscription at this time. The "terminated" value indicates that the subscription is not active.

If the value of the "Subscription-State" header field is "active" or "pending", the notifier MUST also include in the "Subscription-State" header field an "expires" parameter which indicates the time remaining on the subscription. The notifier MAY use this mechanism to shorten a subscription; however, this mechanism MUST NOT be used to lengthen a subscription.

Including expiration information for active and pending subscriptions is necessary in case the SUBSCRIBE request forks, since the response to a forked SUBSCRIBE request may not be received by the subscriber. [RFC3265] allowed the notifier some discretion in the inclusion of this parameter, so subscriber implementations are warned to handle the lack of an "expires" parameter gracefully. Note well that this "expires" value is a parameter on the "Subscription-State" header field, NOT an "Expires" header field.
The period of time for a subscription can be shortened to zero by the notifier. In other words, it is perfectly valid for a SUBSCRIBE request with a non-zero expires to be answered with a NOTIFY request that contains "Subscription-Status: terminated;reason=expired". This merely means that the notifier has shortened the subscription timeout to zero, and the subscription has expired instantaneously. The body may contain valid state, or it may contain a neutral state (see Section 5.4.7).

If the value of the "Subscription-State" header field is "terminated", the notifier SHOULD also include a "reason" parameter. The notifier MAY also include a "retry-after" parameter, where appropriate. For details on the value and semantics of the "reason" and "retry-after" parameters, see Section 4.1.3.

4.2.3. PINT Compatibility

The "Event" header field is considered mandatory for the purposes of this document. However, to maintain compatibility with PINT (see [RFC2848]), notifiers MAY interpret a SUBSCRIBE request with no "Event" header field as requesting a subscription to PINT events. If a notifier does not support PINT, it SHOULD return "489 Bad Event" to any SUBSCRIBE requests without an "Event" header field.

4.3. Proxy Behavior

Proxies need no additional behavior beyond that described in [RFC3261] to support SUBSCRIBE and NOTIFY transactions. If a proxy wishes to see all of the SUBSCRIBE and NOTIFY requests for a given dialog, it MUST add a Record-Route header field to the initial SUBSCRIBE request and all NOTIFY requests. It MAY choose to include Record-Route in subsequent SUBSCRIBE requests; however, these requests cannot cause the dialog’s route set to be modified.

Proxies that did not add a Record-Route header field to the initial SUBSCRIBE request MUST NOT add a Record-Route header field to any of the associated NOTIFY requests.

Note that subscribers and notifiers may elect to use S/MIME encryption of SUBSCRIBE and NOTIFY requests; consequently, proxies cannot rely on being able to access any information that is not explicitly required to be proxy-readable by [RFC3261].

4.4. Common Behavior
4.4.1. Dialog Creation and Termination

Dialogs usages are created upon completion of a NOTIFY transaction for a new subscription, unless the NOTIFY request contains a "Subscription-State" of "terminated."

Because the dialog usage is established by the NOTIFY request, the route set at the subscriber is taken from the NOTIFY request itself, as opposed to the route set present in the 200-class response to the SUBSCRIBE request.

NOTIFY requests are matched to such SUBSCRIBE requests if they contain the same "Call-ID", a "To" header field "tag" parameter which matches the "From" header field "tag" parameter of the SUBSCRIBE request, and the same "Event" header field. Rules for comparisons of the "Event" header fields are described in Section 8.2.1.

A subscription is destroyed after a notifier sends a NOTIFY request with a "Subscription-State" of "terminated," or in certain error situations described elsewhere in this document. The subscriber will generally answer such final requests with a "200 OK" response (unless a condition warranting an alternate response has arisen). Except when the mechanism described in Section 4.5.2 is used, the destruction of a subscription results in the termination of its associated dialog.

A subscriber may send a SUBSCRIBE request with an "Expires" header field of 0 in order to trigger the sending of such a NOTIFY request; however, for the purposes of subscription and dialog lifetime, the subscription is not considered terminated until the NOTIFY transaction with a "Subscription-State" of "terminated" completes.

4.4.2. Notifier Migration

It is often useful to allow migration of subscriptions between notifiers. Such migration may be effected by sending a NOTIFY request with a "Subscription-State" header field of "terminated", and a reason parameter of "deactivated". This NOTIFY request is otherwise normal, and is formed as described in Section 4.2.2.

Upon receipt of this NOTIFY request, the subscriber SHOULD attempt to re-subscribe (as described in the preceding sections). Note that this subscription is established on a new dialog, and does not re-use the route set from the previous subscription dialog.

The actual migration is effected by making a change to the policy (such as routing decisions) of one or more servers to which the
SUBSCRIBE request will be sent in such a way that a different node ends up responding to the SUBSCRIBE request. This may be as simple as a change in the local policy in the notifier from which the subscription is migrating so that it serves as a proxy or redirect server instead of a notifier.

Whether, when, and why to perform notifier migrations may be described in individual event packages; otherwise, such decisions are a matter of local notifier policy, and are left up to individual implementations.

4.4.3. Polling Resource State

A natural consequence of the behavior described in the preceding sections is that an immediate fetch without a persistent subscription may be effected by sending a SUBSCRIBE with an "Expires" of 0.

Of course, an immediate fetch while a subscription is active may be effected by sending a SUBSCRIBE request with an "Expires" equal to the number of seconds remaining in the subscription.

Upon receipt of this SUBSCRIBE request, the notifier (or notifiers, if the SUBSCRIBE request was forked) will send a NOTIFY request containing resource state in the same dialog.

Note that the NOTIFY requests triggered by SUBSCRIBE requests with "Expires" header fields of 0 will contain a "Subscription-State" value of "terminated", and a "reason" parameter of "timeout".

Polling of event state can cause significant increases in load on the network and notifiers; as such, it should be used only sparingly. In particular, polling SHOULD NOT be used in circumstances in which it will typically result in more network messages than long-running subscriptions.

When polling is used, subscribers SHOULD attempt to cache authentication credentials between polls so as to reduce the number of messages sent.

Due to the requirement on notifiers to send a NOTIFY request immediately upon receipt of a SUBSCRIBE request, the state provided by polling is limited to the information that the notifier has immediate local access to when it receives the SUBSCRIBE request. If, for example, the notifier generally needs to retrieve state from another network server, then that state will be absent from the NOTIFY request that results from polling.
4.4.4. Allow-Events header field usage

The "Allow-Events" header field, if present, MUST include a comprehensive and inclusive list of tokens which indicates the event packages for which the User Agent can act as a notifier. In other words, a user agent sending an "Allow-Events" header field is advertising that it can process SUBSCRIBE requests and generate NOTIFY requests for all of the event packages listed in that header field.

Any user agent that can act as a notifier for one or more event packages SHOULD include an appropriate "Allow-Events" header field indicating all supported events in all methods which initiate dialogs and their responses (such as INVITE) and OPTIONS responses.

This information is very useful, for example, in allowing user agents to render particular interface elements appropriately according to whether the events required to implement the features they represent are supported by the appropriate nodes. On the other hand, it doesn’t necessarily make much sense to indicate supported events inside a dialog established by a NOTIFY request if the only event package supported is the one associated with that subscription.

Note that "Allow-Events" header fields MUST NOT be inserted by proxies.

The "Allow-Events" header field does not include a list of the event template packages supported by an implementation. If a subscriber wishes to determine which event template packages are supported by a notifier, it can probe for such support by attempting to subscribe to the event template packages it wishes to use.

For example: to check for support for the templatized package "presence.winfo", a client may attempt to subscribe to that event package for a known resource, using an "Expires" header value of 0. If the response is a 489 error code, then the client can deduce that "presence.winfo" is unsupported.

4.5. Targeting Subscriptions at Devices

[RFC3265] defined a mechanism by which subscriptions could share dialogs with invite usages and with other subscriptions. The purpose of this behavior was to allow subscribers to ensure that a subscription arrived at the same device as an established dialog. Unfortunately, the re-use of dialogs has proven to be exceedingly confusing. [RFC5057] attempted to clarify proper behavior in a variety of circumstances; however, the ensuing rules remain confusing...
and prone to implementation error. At the same time, the mechanism described in [RFC5627] now provides a far more elegant and unambiguous means to achieve the same goal.

Consequently, the dialog re-use technique described in RFC 3265 is now deprecated.

This dialog-sharing technique has also historically been used as a means for targeting an event package at a dialog. This usage can be seen, for example, in certain applications of the REFER method [RFC3515]. With the removal of dialog re-use, an alternate (and more explicit) means of targeting dialogs needs to be used for this type of correlation. The appropriate means of such targeting is left up to the actual event packages. Candidates include the "Target-Dialog" header field [RFC4538], the "Join" header field [RFC3911], and the "Replaces" header field [RFC3891], depending on the semantics desired. Alternately, if the semantics of those header fields do not match the event package’s purpose for correlation, event packages can devise their own means of identifying dialogs. For an example of this approach, see the Dialog Event Package [RFC4235].

4.5.1. Using GRUUs to Route to Devices

Notifiers MUST implement the Globally Routable User-Agent URI (GRUU) extension defined in [RFC5627], and MUST use a GRUU as their local target. This allows subscribers to explicitly target desired devices.

If a subscriber wishes to subscribe to a resource on the same device as an established dialog, it should check whether the remote contact in that dialog is a GRUU (i.e., whether it contains a "gr" URI parameter). If so, the subscriber creates a new dialog, using the GRUU as the request URI for the new SUBSCRIBE request.

Because GRUUs are guaranteed to route to a a specific device, this ensures that the subscription will be routed to the same place as the established dialog.

4.5.2. Sharing Dialogs

For compatibility with older clients, subscriber and notifier implementations may choose to allow dialog sharing. The behavior of multiple usages within a dialog are described in [RFC5057].

Subscribers MUST NOT attempt to re-use dialogs whose remote target is a GRUU.
Note that the techniques described in this section are included for backwards compatibility purposes only. Because subscribers cannot re-use dialogs with a GRUU for their remote target, and because notifiers must use GRUUs as their local target, any two implementations that conform to this specification will automatically use the mechanism described in Section 4.5.1.

Further note that the prohibition on re-using dialogs does not exempt implicit subscriptions created by the REFER method. This means that implementations complying with this specification are required to use the "Target-Dialog" mechanism described in [RFC4538] when the remote target is a GRUU.

If a subscriber wishes to subscribe to a resource on the same device as an established dialog and the remote contact is not a GRUU, it MAY revert to dialog sharing behavior. Alternately, it MAY choose to treat the remote party as incapable of servicing the subscription (i.e., the same way it would behave if the remote party did not support SIP events at all).

If a notifier receives a SUBSCRIBE request for a new subscription on an existing dialog, it MAY choose to implement dialog sharing behavior. Alternately, it may choose to fail the SUBSCRIBE request with a 403 response. The error text of such 403 responses SHOULD indicate that dialog sharing is not supported.

To implement dialog sharing, subscribers and notifiers perform the following additional processing:

- When subscriptions exist in dialogs associated with INVITE-created application state and/or other subscriptions, these sets of application state do not interact beyond the behavior described for a dialog (e.g., route set handling). In particular, multiple subscriptions within a dialog are expire independently, and require independent subscription refreshes.

- If a subscription’s destruction leaves no other application state associated with the dialog, the dialog terminates. The destruction of other application state (such as that created by an INVITE) will not terminate the dialog if a subscription is still associated with that dialog. This means that, when dialog are reused, then a dialog created with an INVITE does not necessarily terminate upon receipt of a BYE. Similarly, in the case that several subscriptions are associated with a single dialog, the dialog does not terminate until all the subscriptions in it are destroyed.
Subscribers MAY include an "id" parameter in SUBSCRIBE request "Event" header field to allow differentiation between multiple subscriptions in the same dialog. This "id" parameter, if present, contains an opaque token which identifies the specific subscription within a dialog. An "id" parameter is only valid within the scope of a single dialog.

If an "id" parameter is present in the SUBSCRIBE request used to establish a subscription, that "id" parameter MUST also be present in all corresponding NOTIFY requests.

When a subscriber refreshes a the subscription timer, the SUBSCRIBE request MUST contain the same "Event" header field "id" parameter as was present in the SUBSCRIBE request that created the subscription. (Otherwise, the notifier will interpret the SUBSCRIBE request as a request for a new subscription in the same dialog).

When a subscription is created in the notifier, it stores any "Event" header field "id" parameter as part of the subscription information (along with the event package name).

If an initial SUBSCRIBE request is sent on a pre-existing dialog, a matching NOTIFY request merely creates a new subscription associated with that dialog.

4.6. CANCEL Requests for SUBSCRIBE and NOTIFY Transactions

Neither SUBSCRIBE nor NOTIFY requests can be canceled. If a UAS receives a CANCEL request that matches a known SUBSCRIBE or NOTIFY transaction, it MUST respond to the CANCEL request, but otherwise ignore it. In particular, the CANCEL request MUST NOT affect processing of the SUBSCRIBE or NOTIFY request in any way.

UACs SHOULD NOT send CANCEL requests for SUBSCRIBE or NOTIFY transactions.

5. Event Packages

This section covers several issues which should be taken into consideration when event packages based on the SUBSCRIBE and NOTIFY methods are proposed.
5.1. Appropriateness of Usage

When designing an event package using the methods described in this document for event notification, it is important to consider: is SIP an appropriate mechanism for the problem set? Is SIP being selected because of some unique feature provided by the protocol (e.g., user mobility), or merely because "it can be done?" If you find yourself defining event packages for notifications related to, for example, network management or the temperature inside your car’s engine, you may want to reconsider your selection of protocols.

Those interested in extending the mechanism defined in this document are urged to follow the development of "Guidelines for Authors of SIP Extensions" [RFC4485] for further guidance regarding appropriate uses of SIP.

Further, it is expected that this mechanism is not to be used in applications where the frequency of reportable events is excessively rapid (e.g., more than about once per second). A SIP network is generally going to be provisioned for a reasonable signaling volume; sending a notification every time a user’s GPS position changes by one hundredth of a second could easily overload such a network.

5.2. Event Template-packages

Normal event packages define a set of state applied to a specific type of resource, such as user presence, call state, and messaging mailbox state.

Event template-packages are a special type of package which define a set of state applied to other packages, such as statistics, access policy, and subscriber lists. Event template-packages may even be applied to other event template-packages.

To extend the object-oriented analogy made earlier, event template-packages can be thought of as templatized C++ packages which must be applied to other packages to be useful.

The name of an event template-package as applied to a package is formed by appending a period followed by the event template-package name to the end of the package. For example, if a template-package called "winfo" were being applied to a package called "presence", the event token used in the "Event" header field would be "presence.winfo".
This scheme may be arbitrarily extended. For example, application of the "winfo" package to the "presence.winfo" state of a resource would be represented by the name "presence.winfo.winfo". It naturally follows from this syntax that the order in which templates are specified is significant.

For example: consider a theoretical event template-package called "list". The event "presence.winfo.list" would be the application of the "list" template to "presence.winfo", which would presumably be a list of winfo state associated with presence. On the other hand, the event "presence.list.winfo" would represent the application of winfo to "presence.list", which would represent the winfo state of a list of presence information.

Event template-packages must be defined so that they can be applied to any arbitrary package. In other words, event template-packages cannot be specifically tied to one or a few "parent" packages in such a way that they will not work with other packages.

5.3. Amount of State to be Conveyed

When designing event packages, it is important to consider the type of information which will be conveyed during a notification.

A natural temptation is to convey merely the event (e.g., "a new voice message just arrived") without accompanying state (e.g., "7 total voice messages"). This complicates implementation of subscribing entities (since they have to maintain complete state for the entity to which they have subscribed), and also is particularly susceptible to synchronization problems.

There are two possible solutions to this problem that event packages may choose to implement.

5.3.1. Complete State Information

In general, event packages need to be able to convey a well-defined and complete state, rather than just a stream of events. If it is not possible to describe complete system state for transmission in NOTIFY requests, then the problem set is not a good candidate for an event package.

For packages which typically convey state information that is reasonably small (on the order of 1 KB or so), it is suggested that event packages are designed so as to send complete state information whenever an event occurs.

In some circumstances, conveying the current state alone may be
insufficient for a particular class of events. In these cases, the event packages should include complete state information along with the event that occurred. For example, conveying "no customer service representatives available" may not be as useful as conveying "no customer service representatives available; representative sip:46@cs.xyz.int just logged off".

5.3.2. State Deltas

In the case that the state information to be conveyed is large, the event package may choose to detail a scheme by which NOTIFY requests contain state deltas instead of complete state.

Such a scheme would work as follows: any NOTIFY request sent in immediate response to a SUBSCRIBE request contains full state information. NOTIFY requests sent because of a state change will contain only the state information that has changed; the subscriber will then merge this information into its current knowledge about the state of the resource.

Any event package that supports delta changes to states MUST include a version number that increases by exactly one for each NOTIFY transaction in a subscription. Note that the state version number appears in the body of the message, not in a SIP header field.

If a NOTIFY request arrives that has a version number that is incremented by more than one, the subscriber knows that a state delta has been missed; it ignores the NOTIFY request containing the state delta (except for the version number, which it retains to detect message loss), and re-sends a SUBSCRIBE request to force a NOTIFY request containing a complete state snapshot.

5.4. Event Package Responsibilities

Event packages are not required to reiterate any of the behavior described in this document, although they may choose to do so for clarity or emphasis. In general, though, such packages are expected to describe only the behavior that extends or modifies the behavior described in this document.

Note that any behavior designated with "SHOULD" or "MUST" in this document is not allowed to be weakened by extension documents; however, such documents may elect to strengthen "SHOULD" requirements to "MUST" strength if required by their application.

In addition to the normal sections expected in standards-track RFCs and SIP extension documents, authors of event packages need to address each of the issues detailed in the following subsections.
For clarity: well-formed event package definitions contain sections addressing each of these issues, ideally in the same order and with the same titles as these subsections.

5.4.1. Event Package Name

This section, which MUST be present, defines the token name to be used to designate the event package. It MUST include the information which appears in the IANA registration of the token. For information on registering such types, see Section 7.

5.4.2. Event Package Parameters

If parameters are to be used on the "Event" header field to modify the behavior of the event package, the syntax and semantics of such header fields MUST be clearly defined.

Any "Event" header field parameters defined by an event package MUST be registered in the "Header Field Parameters and Parameter Values" registry defined by [RFC3968]. An "Event" header field parameter, once registered in conjunction with an event package, MUST NOT be re-used with any other event package. Non-event-package specifications MAY define "Event" header field parameters that apply across all event packages (with emphasis on "all", as opposed to "several"), such as the "id" parameter defined in this document. The restriction of a parameter to use with a single event package only applies to parameters that are defined in conjunction with an event package.

5.4.3. SUBSCRIBE Request Bodies

It is expected that most, but not all, event packages will define syntax and semantics for SUBSCRIBE request bodies; these bodies will typically modify, expand, filter, throttle, and/or set thresholds for the class of events being requested. Designers of event packages are strongly encouraged to re-use existing media types for message bodies where practical. See [RFC4288] for information on media type specification and registration.

This mandatory section of an event package defines what type or types of event bodies are expected in SUBSCRIBE requests (or specify that no event bodies are expected). It should point to detailed definitions of syntax and semantics for all referenced body types.

5.4.4. Subscription Duration

It is RECOMMENDED that event packages give a suggested range of times considered reasonable for the duration of a subscription. Such packages MUST also define a default "Expires" value to be used if
none is specified.

5.4.5. NOTIFY Request Bodies

The NOTIFY request body is used to report state on the resource being monitored. Each package MUST define what type or types of event bodies are expected in NOTIFY requests. Such packages MUST specify or cite detailed specifications for the syntax and semantics associated with such event body.

Event packages also MUST define which media type is to be assumed if none are specified in the "Accept" header field of the SUBSCRIBE request.

5.4.6. Notifier processing of SUBSCRIBE requests

This section describes the processing to be performed by the notifier upon receipt of a SUBSCRIBE request. Such a section is required.

Information in this section includes details of how to authenticate subscribers and authorization issues for the package.

5.4.7. Notifier generation of NOTIFY requests

This section of an event package describes the process by which the notifier generates and sends a NOTIFY request. This includes detailed information about what events cause a NOTIFY request to be sent, how to compute the state information in the NOTIFY, how to generate neutral or fake state information to hide authorization delays and decisions from users, and whether state information is complete or deltas for notifications; see Section 5.3. Such a section is required.

This section may optionally describe the behavior used to process the subsequent response.

5.4.8. Subscriber processing of NOTIFY requests

This section of an event package describes the process followed by the subscriber upon receipt of a NOTIFY request, including any logic required to form a coherent resource state (if applicable).

5.4.9. Handling of forked requests

Each event package MUST specify whether forked SUBSCRIBE requests are allowed to install multiple subscriptions.

If such behavior is not allowed, the first potential dialog-
establishing message will create a dialog. All subsequent NOTIFY requests which correspond to the SUBSCRIBE request (i.e., match "To", "From", "From" header field "tag" parameter, "Call-ID", "Event", and "Event" header field "id" parameter) but which do not match the dialog would be rejected with a 481 response. Note that the 200-class response to the SUBSCRIBE request can arrive after a matching NOTIFY request has been received; such responses might not correlate to the same dialog established by the NOTIFY request. Except as required to complete the SUBSCRIBE transaction, such non-matching 200-class responses are ignored.

If installing of multiple subscriptions by way of a single forked SUBSCRIBE request is allowed, the subscriber establishes a new dialog towards each notifier by returning a 200-class response to each NOTIFY request. Each dialog is then handled as its own entity, and is refreshed independent of the other dialogs.

In the case that multiple subscriptions are allowed, the event package MUST specify whether merging of the notifications to form a single state is required, and how such merging is to be performed. Note that it is possible that some event packages may be defined in such a way that each dialog is tied to a mutually exclusive state which is unaffected by the other dialogs; this MUST be clearly stated if it is the case.

5.4.10. Rate of notifications

Each event package is expected to define a requirement (SHOULD or MUST strength) which defines an absolute maximum on the rate at which notifications are allowed to be generated by a single notifier.

Each package MAY further define a throttle mechanism which allows subscribers to further limit the rate of notification.

5.4.11. State Aggregation

Many event packages inherently work by collecting information about a resource from a number of other sources -- either through the use of PUBLISH [RFC3903], by subscribing to state information, or through other state gathering mechanisms.

Event packages that involve retrieval of state information for a single resource from more than one source need to consider how notifiers aggregate information into a single, coherent state. Such packages MUST specify how notifiers aggregate information and how they provide authentication and authorization.
5.4.12. Examples

Event packages SHOULD include several demonstrative message flow diagrams paired with several typical, syntactically correct, and complete messages.

It is RECOMMENDED that documents describing event packages clearly indicate that such examples are informative and not normative, with instructions that implementors refer to the main text of the document for exact protocol details.

5.4.13. Use of URIs to Retrieve State

Some types of event packages may define state information which is potentially too large to reasonably send in a SIP message. To alleviate this problem, event packages may include the ability to convey a URI instead of state information; this URI will then be used to retrieve the actual state information.

[RFC4483] defines a mechanism that can be used by event packages to convey information in such a fashion.

6. Security Considerations

6.1. Access Control

The ability to accept subscriptions should be under the direct control of the notifier’s user, since many types of events may be considered sensitive for the purposes of privacy. Similarly, the notifier should have the ability to selectively reject subscriptions based on the subscriber identity (based on access control lists), using standard SIP authentication mechanisms. The methods for creation and distribution of such access control lists is outside the scope of this document.

6.2. Notifier Privacy Mechanism

The mere act of returning certain 4xx and 6xx responses to SUBSCRIBE requests may, under certain circumstances, create privacy concerns by revealing sensitive policy information. In these cases, the notifier SHOULD always return a 200 (OK) response. While the subsequent NOTIFY request may not convey true state, it MUST appear to contain a potentially correct piece of data from the point of view of the subscriber, indistinguishable from a valid response. Information about whether a user is authorized to subscribe to the requested state is never conveyed back to the original user under these circumstances.
Individual packages and their related documents for which such a mode of operation makes sense can further describe how and why to generate such potentially correct data. For example, such a mode of operation is mandated by [RFC2779] for user presence information.

6.3. Denial-of-Service attacks

The current model (one SUBSCRIBE request triggers a SUBSCRIBE response and one or more NOTIFY requests) is a classic setup for an amplifier node to be used in a smurf attack.

Also, the creation of state upon receipt of a SUBSCRIBE request can be used by attackers to consume resources on a victim’s machine, rendering it unusable.

To reduce the chances of such an attack, implementations of notifiers SHOULD require authentication. Authentication issues are discussed in [RFC3261].

6.4. Replay Attacks

Replaying of either SUBSCRIBE or NOTIFY requests can have detrimental effects.

In the case of SUBSCRIBE requests, attackers may be able to install any arbitrary subscription which it witnessed being installed at some point in the past. Replaying of NOTIFY requests may be used to spoof old state information (although a good versioning mechanism in the body of the NOTIFY requests may help mitigate such an attack). Note that the prohibition on sending NOTIFY requests to nodes which have not subscribed to an event also aids in mitigating the effects of such an attack.

To prevent such attacks, implementations SHOULD require authentication with anti-replay protection. Authentication issues are discussed in [RFC3261].

6.5. Man-in-the-middle attacks

Even with authentication, man-in-the-middle attacks using SUBSCRIBE requests may be used to install arbitrary subscriptions, hijack existing subscriptions, terminate outstanding subscriptions, or modify the resource to which a subscription is being made. To prevent such attacks, implementations SHOULD provide integrity protection across "Contact", "Route", "Expires", "Event", and "To" header fields of SUBSCRIBE requests, at a minimum. If SUBSCRIBE request bodies are used to define further information about the state of the call, they SHOULD be included in the integrity protection
scheme.

Man-in-the-middle attacks may also attempt to use NOTIFY requests to spoof arbitrary state information and/or terminate outstanding subscriptions. To prevent such attacks, implementations SHOULD provide integrity protection across the "Call-ID", "CSeq", and "Subscription-State" header fields and the bodies of NOTIFY requests.

Integrity protection of message header fields and bodies is discussed in [RFC3261].

6.6. Confidentiality

The state information contained in a NOTIFY request has the potential to contain sensitive information. Implementations MAY encrypt such information to ensure confidentiality.

While less likely, it is also possible that the information contained in a SUBSCRIBE request contains information that users might not want to have revealed. Implementations MAY encrypt such information to ensure confidentiality.

To allow the remote party to hide information it considers sensitive, all implementations SHOULD be able to handle encrypted SUBSCRIBE and NOTIFY requests.

The mechanisms for providing confidentiality are detailed in [RFC3261].

7. IANA Considerations

(This section is not applicable until this document is published as an RFC.)

With the exception of Section 7.2, the subsections here are for current reference, carried over from the original specification. The only IANA actions requested here are updating all registry references that point to RFC 3265 to instead indicate this document, and creating the new "reason code" registry described in Section 7.2.

7.1. Event Packages

This document defines an event-type namespace which requires a central coordinating body. The body chosen for this coordination is the Internet Assigned Numbers Authority (IANA).

There are two different types of event-types: normal event packages,
and event template-packages; see Section 5.2. To avoid confusion, template-package names and package names share the same namespace; in other words, an event template-package are forbidden from sharing a name with a package.

Policies for registration of SIP event packages and SIP event package templates are defined in section 4.1 of [RFC5727].

Registrations with the IANA are required to include the token being registered and whether the token is a package or a template-package. Further, packages must include contact information for the party responsible for the registration and/or a published document which describes the event package. Event template-package token registrations are also required to include a pointer to the published RFC which defines the event template-package.

Registered tokens to designate packages and template-packages are disallowed from containing the character ".", which is used to separate template-packages from packages.

7.1.1. Registration Information

As this document specifies no package or template-package names, the initial IANA registry for event types will be empty. The remainder of the text in this section gives an example of the type of information to be maintained by the IANA; it also demonstrates all five possible permutations of package type, contact, and reference.

The table below lists the event packages and template-packages defined in "SIP-Specific Event Notification" [RFC xxxx]. Each name is designated as a package or a template-package under "Type".

<table>
<thead>
<tr>
<th>Package Name</th>
<th>Type</th>
<th>Contact</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>example1</td>
<td>package</td>
<td>[Roach]</td>
<td>[RFC xxxx]</td>
</tr>
<tr>
<td>example2</td>
<td>package</td>
<td>[Roach]</td>
<td>[RFC xxxx]</td>
</tr>
<tr>
<td>example3</td>
<td>package</td>
<td></td>
<td>[RFC xxxx]</td>
</tr>
<tr>
<td>example4</td>
<td>template</td>
<td>[Roach]</td>
<td>[RFC xxxx]</td>
</tr>
<tr>
<td>example5</td>
<td>template</td>
<td>[Roach]</td>
<td>[RFC xxxx]</td>
</tr>
</tbody>
</table>

PEOPLE
------
[Roach] Adam Roach <adam.roach@tekelec.com>

REFERENCES
----------
[RFC xxxx] A.B. Roach, "SIP-Specific Event Notification", RFC XXXX,
7.1.2. Registration Template

To: ietf-sip-events@iana.org
Subject: Registration of new SIP event package

Package Name:

(Package names must conform to the syntax described in Section 8.2.1.)

Is this registration for a Template Package:

(indicate yes or no)

Published Specification(s):

(Template packages require a published RFC. Other packages may reference a specification when appropriate).

Person & email address to contact for further information:

7.2. Reason Codes

This document further defines "reason" codes for use in the "Subscription-State" header field (see Section 4.1.3).

Following the policies outlined in "Guidelines for Writing an IANA Considerations Section in RFCs" [RFC5226], new reason codes require a Standards Action.

Registrations with the IANA include the reason code being registered and a reference to a published document which describes the event package. Insertion of such values takes place as part of the RFC publication process or as the result of inter-SDO liaison activity, the result of which will be publication of an associated RFC. New reason codes must conform to the syntax of the ABNF "token" element defined in [RFC3261].

[RFC4660] defined a new reason code prior to the establishment of an IANA registry. We include its reason code ("badfilter") in the initial list of reason codes to ensure a complete registry.

The IANA registry for reason code will be initialized with the following values:
<table>
<thead>
<tr>
<th>Reason Code</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>deactivated</td>
<td>[RFC xxxx]</td>
</tr>
<tr>
<td>probation</td>
<td>[RFC xxxx]</td>
</tr>
<tr>
<td>rejected</td>
<td>[RFC xxxx]</td>
</tr>
<tr>
<td>timeout</td>
<td>[RFC xxxx]</td>
</tr>
<tr>
<td>giveup</td>
<td>[RFC xxxx]</td>
</tr>
<tr>
<td>noresource</td>
<td>[RFC xxxx]</td>
</tr>
<tr>
<td>invariant</td>
<td>[RFC xxxx]</td>
</tr>
<tr>
<td>badfilter</td>
<td>[RFC 4660]</td>
</tr>
</tbody>
</table>

REFERENCES
----------


7.3. Header Field Names

This document registers three new header field names, described elsewhere in this document. These header fields are defined by the following information, which is to be added to the header field sub-registry under http://www.iana.org/assignments/sip-parameters.

- Header Name: Allow-Events
  Compact Form: u

- Header Name: Subscription-State
  Compact Form: (none)

- Header Name: Event
  Compact Form: o

7.4. Response Codes

This document registers two new response codes. These response codes are defined by the following information, which is to be added to the method and response-code sub-registry under http://www.iana.org/assignments/sip-parameters.

- Response Code Number: 202
  Default Reason Phrase: Accepted

- Response Code Number: 489
  Default Reason Phrase: Bad Event
8. Syntax

This section describes the syntax extensions required for event notification in SIP. Semantics are described in Section 4. Note that the formal syntax definitions described in this document are expressed in the ABNF format used in [RFC3261], and contain references to elements defined therein.

8.1. New Methods

This document describes two new SIP methods: SUBSCRIBE and NOTIFY.

8.1.1. SUBSCRIBE method

"SUBSCRIBE" is added to the definition of the element "Method" in the SIP message grammar.

Like all SIP method names, the SUBSCRIBE method name is case sensitive. The SUBSCRIBE method is used to request asynchronous notification of an event or set of events at a later time.

8.1.2. NOTIFY method

"NOTIFY" is added to the definition of the element "Method" in the SIP message grammar.

The NOTIFY method is used to notify a SIP node that an event which has been requested by an earlier SUBSCRIBE method has occurred. It may also provide further details about the event.

8.2. New Header Fields

8.2.1. "Event" Header Field

Event is added to the definition of the element "message-header field" in the SIP message grammar.

For the purposes of matching NOTIFY requests with SUBSCRIBE requests, the event-type portion of the "Event" header field is compared byte-by-byte, and the "id" parameter token (if present) is compared byte-by-byte. An "Event" header field containing an "id" parameter never matches an "Event" header field without an "id" parameter. No other parameters are considered when performing a comparison. SUBSCRIBE responses are matched per the transaction handling rules in [RFC3261].
Note that the forgoing text means that "Event: foo; id=1234" would match "Event: foo; param=abcd; id=1234", but not "Event: foo" (id does not match) or "Event: Foo; id=1234" (event portion does not match).

This document does not define values for event-types. These values will be defined by individual event packages, and MUST be registered with the IANA.

There MUST be exactly one event type listed per event header field. Multiple events per message are disallowed.

The "Event" header field is defined only for use in SUBSCRIBE and NOTIFY requests, and other requests whose definition explicitly calls for its use. It MUST NOT appear in any other SIP requests, and MUST NOT appear in responses.

8.2.2. "Allow-Events" Header Field

Allow-Events is added to the definition of the element "general-header field" in the SIP message grammar. Its usage is described in Section 4.4.4.

User Agents MAY include the "Allow-Events" header field in any request or response, as long as its contents comply with the behavior described in Section 4.4.4.

8.2.3. "Subscription-State" Header Field

Subscription-State is added to the definition of the element "request-header field" in the SIP message grammar. Its usage is described in Section 4.1.3. "Subscription-State" header fields are defined for use in NOTIFY requests only. They MUST NOT appear in other SIP requests or responses.

8.3. New Response Codes

8.3.1. "202 Accepted" Response Code

For historical purposes, the 202 (Accepted) response code is added to the "Success" header field definition.

This document does not specify the use of the 202 response code in conjunction with the SUBSCRIBE or NOTIFY methods. Previous versions of the SIP Events Framework assigned specific meaning to the 202 response code.

Due to response handling in forking cases, any 202 response to a
SUBSCRIBE request may be absorbed by a proxy, and thus it can never be guaranteed to be received by the UAC. Furthermore, there is no actual processing difference for a 202 as compared to a 200; a NOTIFY request is sent after the subscription is processed, and it conveys the correct state. SIP interoperability tests found that implementations were handling 202 differently from 200, leading to incompatibilities. Therefore, the 202 response is being deprecated to make it clear there is no such difference and 202 should not be handled differently than 200.

Implementations conformant with the current specification MUST treat an incoming 202 response as identical to a 200 response, and MUST NOT generate 202 response codes to SUBSCRIBE or NOTIFY requests.

This document also updates [RFC4660], which reiterates the 202-based behavior in several places. Implementations compliant with the present document MUST NOT send a 202 response to a SUBSCRIBE request, and will send an alternate success response (such as 200) in its stead.

8.3.2. "489 Bad Event" Response Code

The 489 event response is added to the "Client-Error" header field definition. "489 Bad Event" is used to indicate that the server did not understand the event package specified in a "Event" header field.

8.4. Augmented BNF Definitions

The Augmented BNF definitions for the various new and modified syntax elements follows. The notation is as used in [RFC3261], and any elements not defined in this section are as defined in SIP and the documents to which it refers.
SUBSCRIBE = %x53.55.42.53.43.52.49.42.45 ; SUBSCRIBE in caps
NOTIFYm = %x4E.4F.54.49.46.59 ; NOTIFY in caps
extension-method = SUBSCRIBE / NOTIFYm / token

Event = ( "Event" / "o" ) HCOLON event-type *( SEMI event-param )
event-type = event-package *( "." event-template )
event-package = token-nodot
event-template = token-nodot
token-nodot = 1*( alphanum / "-" / "!" / "#" / "*" / "/" / "_" / "^" / "" / "\" )

; The use of the "id" parameter is deprecated; it is included
; for backwards compatibility purposes only.
event-param = generic-param / ( "id" EQUAL token )

Allow-Events = ( "Allow-Events" / "u" ) HCOLON event-type *(COMMA event-type)

Subscription-State = "Subscription-State" HCOLON substate-value *( SEMI subexp-params )
substate-value = "active" / "pending" / "terminated"
extension-substate = token
subexp-params = ("reason" EQUAL event-reason-value)
/ ("expires" EQUAL delta-seconds)
/ ("retry-after" EQUAL delta-seconds)
/ generic-param
event-reason-value = "deactivated"
/ "probation"
/ "rejected"
/ "timeout"
/ "giveup"
/ "noresource"
/ "invariant"
/ event-reason-extension

event-reason-extension = token

9. References

9.1. Normative References


9.2. Informative References


Appendix A. Acknowledgements

Thanks to the participants in the Events BOF at the 48th IETF meeting in Pittsburgh, as well as those who gave ideas and suggestions on the SIP Events mailing list. In particular, I wish to thank Henning Schulzrinne of Columbia University for coming up with the final three-tiered event identification scheme, Sean Olson for miscellaneous guidance, Jonathan Rosenberg for a thorough scrubbing of the -00 draft, and the authors of the "SIP Extensions for Presence" document for their input to SUBSCRIBE and NOTIFY request semantics.

I also owe a debt of gratitude to all the implementors who have provided feedback on areas of confusion or difficulty in the original specification. In particular, Robert Sparks’ Herculean efforts...
organizing, running, and collecting data from the SIPit events have proven invaluable in shaking out specification bugs. Robert Sparks is also responsible for untangling the dialog usage mess, in the form of RFC 5057 [RFC5057].

Appendix B. Changes from RFC 3265

This document represents several changes from the mechanism originally described in RFC 3265. This section summarizes those changes. Bug numbers refer to the identifiers for the bug reports kept on file at http://bugs.sipit.net/.

B.1. Bug 666: Clarify use of expires=xxx with terminated

Strengthened language in Section 4.1.3 to clarify that expires should not be sent with terminated, and must be ignored if received.

B.2. Bug 667: Reason code for unsub/poll not clearly spelled out

Clarified description of "timeout" in Section 4.1.3. (n.b., the text in Section 4.4.3 is actually pretty clear about this).

B.3. Bug 669: Clarify: SUBSCRIBE for a duration might be answered with a NOTIFY/expires=0

Added clarifying text to Section 4.2.2 explaining that shortening a subscription to zero seconds is valid. Also added sentence to Section 3.1.1 explicitly allowing shortening to zero.

B.4. Bug 670: Dialog State Machine needs clarification

The issues associated with the bug deal exclusively with the handling of multiple usages with a dialog. This behavior has been deprecated and moved to Section 4.5.2. This section, in turn, cites [RFC5057], which addresses all of the issues in Bug 670.

B.5. Bug 671: Clarify timeout-based removal of subscriptions

Changed Section 4.2.2 to specifically cite Timer F (so as to avoid ambiguity between transaction timeouts and retransmission timeouts).

B.6. Bug 672: Mandate expires= in NOTIFY

Changed strength of including of "expires" in a NOTIFY from SHOULD to MUST in Section 4.2.2.
B.7. Bug 673: INVITE 481 response effect clarification

This bug was addressed in [RFC5057].

B.8. Bug 677: SUBSCRIBE response matching text in error

Fixed Section 8.2.1 to remove incorrect "...responses and..." -- explicitly pointed to SIP for transaction response handling.

B.9. Bug 695: Document is not explicit about response to NOTIFY at subscription termination

Added text to Section 4.4.1 indicating that the typical response to a terminal NOTIFY is a "200 OK".

B.10. Bug 696: Subscription state machine needs clarification

Added state machine diagram to Section 4.1.2 with explicit handling of what to do when a SUBSCRIBE never shows up. Added definition of and handling for new Timer N to Section 4.1.2.4. Added state machine to Section 4.2.2 to reinforce text.

B.11. Bug 697: Unsubscription behavior could be clarified

Added text to Section 4.2.1.4 encouraging (but not requiring) full state in final NOTIFY request. Also added text to Section 4.1.2.3 warning subscribers that full state may or may not be present in the final NOTIFY.

B.12. Bug 699: NOTIFY and SUBSCRIBE are target refresh requests

Added text to both Section 3.1 and Section 3.2 explicitly indicating that SUBSCRIBE and NOTIFY are target refresh methods.

B.13. Bug 722: Inconsistent 423 reason phrase text

Changed reason code to "Interval Too Brief" in Section 4.2.1.1 and Section 4.2.1.4, to match 423 reason code in SIP [RFC3261].

B.14. Bug 741: guidance needed on when to not include Allow-Events

Added non-normative clarification to Section 4.4.4 regarding inclusion of Allow-Events in a NOTIFY for the one-and-only package supported by the notifier.
B.15. Bug 744: 5xx to NOTIFY terminates a subscription, but should not

Issue of subscription (usage) termination versus dialog termination is handled in [RFC5057]. The text in Section 4.2.2 has been updated to summarize the behavior described by 5057, and cites it for additional detail and rationale.

B.16. Bug 752: Detection of forked requests is incorrect

Removed erroneous "CSeq" from list of matching criteria in Section 5.4.9.

B.17. Bug 773: Reason code needs IANA registry

Added Section 7.2 to create and populate IANA registry.

B.18. Bug 774: Need new reason for terminating subscriptions to resources that never change

Added new "invariant" reason code to Section 4.1.3, ABNF syntax.

B.19. Clarify handling of Route/Record-Route in NOTIFY

Changed text in Section 4.3 mandating Record-Route in initial SUBSCRIBE and all NOTIFY requests, and adding "MAY" level statements for subsequent SUBSCRIBE requests.

B.20. Eliminate implicit subscriptions

Added text to Section 4.2.1 explaining some of the problems associated with implicit subscriptions, normative language prohibiting them. Removed language from Section 3.2 describing "non-SUBSCRIBE" mechanisms for creating subscriptions. Simplified language in Section 4.2.2, now that the soft-state/non-soft-state distinction is unnecessary.

B.21. Deprecate dialog re-use

Moved handling of dialog re-use and "id" handling to Section 4.5.2. It is documented only for backwards-compatibility purposes.

B.22. Rationalize dialog creation

Section 4.4.1 has been updated to specify that dialogs should be created when the NOTIFY arrives. Previously, the dialog was established by the SUBSCRIBE 200, or by the NOTIFY transaction. This was unnecessarily complicated; the newer rules are easier to implement (and result in effectively the same behavior on the wire).
B.23. Refactor behavior sections

Reorganized Section 4 to consolidate behavior along role lines (subscriber/notifier/proxy) instead of method lines.

B.24. Clarify sections that need to be present in event packages

Added sentence to Section 5 clarifying that event packages are expected to include explicit sections covering the issues discussed in this section.

B.25. Make CANCEL handling more explicit

Text in Section 4.6 now clearly calls out behavior upon receipt of a CANCEL. We also echo the "...SHOULD NOT send..." requirement from [RFC3261].

B.26. Remove State Agent Terminology

As originally planned, we anticipated a fairly large number of event packages that would move back and forth between end-user devices and servers in the network. In practice, this has ended up not being the case. Certain events, like dialog state, are inherently hosted at end-user devices; others, like presence, are almost always hosted in the network (due to issues like composition, and the ability to deliver information when user devices are offline). Further, the concept of State Agents is the most misunderstood by event package authors. In my expert review of event packages, I have yet to find one that got the concept of State Agents completely correct -- and most of them start out with the concept being 100% backwards from the way RFC 3265 described it.

Rather than remove the ability to perform the actions previously attributed to the widely misunderstood term "State Agent," we have simply eliminated this term. Instead, we talk about the behaviors required to create state agents (state aggregation, subscription notification) without defining a formal term to describe the servers that exhibit these behaviors. In effect, this is an editorial change to make life easier for event package authors; the actual protocol does not change as a result.

The definition of "State Agent" has been removed from Section 2. Section 4.4.2 has been retooled to discuss migration of subscription in general, without calling out the specific example of state agents. Section 5.4.11 has been focused on state aggregation in particular, instead of state aggregation as an aspect of state agents.
B.27. Miscellaneous Changes

The following changes are relatively minor revisions to the document that resulted primarily from review of this document in the working group and IESG, rather than implementation reports.

- Clarified scope of Event header field parameters. In RFC3265, the scope is ambiguous, which causes problems with the RFC3968 registry. The new text ensures that Event header field parameters are unique across all event packages.

- Removed obsoleted language around IANA registration policies for event packages. Instead, we now cite RFC5727, which supersedes RFC3265, and is authoritative on event package registration policy.

- Several editorial updates after input from working group, including proper designation of "dialog usage" rather than "dialog" where appropriate.

- Clarified two normative statements about subscription termination by changing from plain English prose to RFC2119 language.

- Removed "Table 2" expansions, per WG consensus on how SIP table 2 is to be handled.

- Removed 202 response code.

- Clarified that "Allow-Events" does not list event template packages.

- Added clarification about proper response when the SUBSCRIBE indicates an unknown media type in its Accept header field.

- Minor clarifications to Route and Record-Route behavior.

- Added non-normative warning about the limitations of state polling.

- Added information about targeting subscriptions at specific dialogs.

- Added RFC 3261 to list of documents updated by this one (rather than the "2543" indicated by RFC3265).

- Clarified text in Section 3.1.1 explaining the meaning of "Expires: 0".
o Changed text in definition of "probation" reason code to indicate that subscribers don’t need to re-subscribe if the associated state is no longer of use to them.

o Specified that the termination of a subscription due to a NOTIFY transaction failure does not require sending another NOTIFY message.

o Clarified how order of template application affects the meaning of an Event header field value. (e.g., "foo.bar.baz" is different than "foo.baz.bar").

Author’s Address

Adam Roach  
Tekelec  
17210 Campbell Rd.  
Suite 250  
Dallas, TX  75252  
US  

Email: adam@nostrum.com
Abstract

This document defines a standard mechanism for capturing the history information associated with a Session Initiation Protocol (SIP) request. This capability enables many enhanced services by providing the information as to how and why a call arrives at a specific application or user. This document defines an optional SIP header, History-Info, for capturing the history information in requests. SIP header field parameters are defined to tag the method by which the target of a request is determined. In addition, this document defines a value for the Privacy header specific to the History-Info header.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 29, 2011.

Copyright Notice
1. Introduction

Many services that SIP is anticipated to support require the ability to determine why and how the call arrived at a specific application. Examples of such services include (but are not limited to) sessions initiated to call centers via "click to talk" SIP Uniform Resource Locators (URLs) on a web page, "call history/logging" style services within intelligent "call management" software for SIP User Agents (UAs), and calls to voicemail servers. Although SIP implicitly provides the retarget capabilities that enable calls to be routed to chosen applications, there is a need for a standard mechanism within SIP for communicating the retargeting history of such a request. This "request history" information allows the receiving application to determine hints about how and why the call arrived at the application/user.

This document defines a SIP header, History-Info, to provide a standard mechanism for capturing the request history information to enable a wide variety of services for networks and end-users. SIP header field parameters are defined to tag the method by which the target of a request is determined. In addition, this document defines a value for the Privacy header specific to the History-Info header.

The History-Info header provides a building block for development of SIP based applications and services. The requirements for the solution described in this document are included in Appendix A. Example scenarios using the History-Info header are included in Appendix B.

2. Conventions and Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

The term "retarget" is used in this document to refer to the process of a SIP entity changing a Uniform Resource Identifier (URI) in a request based on the rules for determining request targets as described in Section 16.5 of [RFC3261] and of the subsequent forwarding of that request as described in step 2 in section 16.6 of [RFC3261]. This includes changing the Request-URI due to a location service lookup and redirect processing. This also includes internal (to a proxy/SIP intermediary) changes of the URI prior to forwarding of the request.

The terms "location service", "forward", "redirect" and "AOR" are
used consistent with the terminology in [RFC3261].

The references to "domain for which the SIP entity/Proxy/Intermediary is responsible" are consistent with and intended to convey the same context as the usage of that terminology in [RFC3261]. The applicability of History-Info to architectures or models outside the context of [RFC3261] is outside the scope of this specification.

3. Background

SIP implicitly provides retargeting capabilities that enable calls to be routed to specific applications as defined in [RFC3261]. The motivation for capturing the request history is that in the process of retargeting a request, old routing information can be forever lost. This lost information may be important history that allows elements to which the call is retargeted to process the call in a locally defined, application-specific manner. This document defines a mechanism for transporting the request history. Application-specific behavior is outside the scope of this specification.

Current network applications provide the ability for elements involved with the call to exchange additional information relating to how and why the call was routed to a particular destination. The following are examples of such applications:

1. Web "referral" applications, whereby an application residing within a web server determines that a visitor to a website has arrived at the site via an "associate" site that will receive some "referral" commission for generating this traffic

2. Email forwarding whereby the forwarded-to user obtains a "history" of who sent the email to whom and at what time

3. Traditional telephony services such as voicemail, call-center "automatic call distribution", and "follow-me" style services

Several of the aforementioned applications currently define application-specific mechanisms through which it is possible to obtain the necessary history information.

In addition, request history information could be used to enhance basic SIP functionality by providing the following:

- Some diagnostic information for debugging SIP requests.
- Capturing aliases and Globally Routable User Agent URIs (GRUUs) [RFC5627], which can be overwritten by a home proxy upon receipt
of the initial request.

- Facilitating the use of limited use addresses (minted on demand) and sub-addressing.
- Preserving service specific URIs that can be overwritten by a downstream proxy, such as those defined in [RFC3087], and control of network announcements and IVR with SIP URI [RFC4240].

4. Overview of Operation

The fundamental functionality provided by the request history information is the ability to inform proxies and UAs involved in processing a request about the history or progress of that request (CAPABILITY-req, see Appendix A). The solution is to capture the Request-URIs as a request is retargeted, in a SIP header: History-Info (CONTENT-req, see Appendix A). This allows for the capturing of the history of a request that would be lost with the normal SIP processing involved in the subsequent retargeting of the request.

The History-Info header can appear in any request not associated with an early or established dialog (e.g., INVITE, REGISTER, MESSAGE, REFER and OPTIONS, PUBLISH and SUBSCRIBE, etc.) (REQUEST-VALIDITY-req, see Appendix A) and any provisional or final responses to these requests (ISSUER-req, see Appendix A).

The following information is carried in the History-Info header as detailed in Section 7.1:

- Targeted-to-URI: The targeted-to-URI entry captures the Request-URI for the specific request as it is forwarded.
- Index: The index reflects the chronological order of the information, indexed to also reflect the forking and nesting of requests.
- Reason: Reason describes why an entry was retargeted.
- Privacy: Privacy is used to request that an entry be anonymized if the request is retargeted to a domain for which the retargeting entity is not responsible.
- Target: A parameter indicating whether the Targeted-to-URI is a registered contact ("rc") for another user mapped ("mp") from the Request-URI in the incoming request that was retargeted. The index of the History-Info entry for the URI that was retargeted is included in each of these parameters. Note that there may be
other reasons a request is retargeted such as normal routing and forwarding, strict routing, etc., thus not all history-info entries have a target header field parameter. The "rc" and "mp" scenarios are what is anticipated to be most useful to end applications/users.

In addition, this specification defines a value for the Privacy header, "history", that applies to all the History-Info header entries in a Request or to a specific History-Info header entry as described above. Further detailed is provided in Section 7.3.1.

The History-Info header is added to a Request when a new request is created by a UAC or forwarded by a Proxy, or when the target of a request is changed. It is possible for the target of a request to be changed by the same proxy/SIP Intermediary multiple times (referred to as "internal retargeting"). A SIP entity changing the target of a request in response to a redirect or REFER also propagates any History-Info header from the initial Request in the new request.

The following is an illustrative example of usage of History-Info.

In this example, Alice (sip:alice@atlanta.example.com) calls Bob (sip:bob@biloxi.example.com). Alice’s home proxy (sip: atlanta.example.com) forwards the request to Bob’s proxy (sip: biloxi.example.com). When the request arrives at sip: biloxi.example.com, it does a location service lookup for bob@biloxi.example.com and changes the target of the request to Bob’s Contact URIs provided as part of normal SIP registration. In this example, Bob is simultaneously contacted on a PC client and on a phone, and Bob answers on the PC client.

One important thing illustrated by this call flow is that without History-Info, Bob would "lose" the target information, including any parameters in the request URI. Bob can recover that information by locating the last hi-entry with an "rc" header field parameter. This "rc" parameter contains the index of the hi-entry containing the lost target information - i.e., the sip:bob@biloxi.example.com entry with index=1.1. Note that an hi-entry is not included for the fork to sip:bob@192.0.2.7 since there was no response at the time the 200 OK is sent to Alice.

The formatting in this scenario is for visual purposes; thus, backslash and CRLF are used between the fields for readability and the headers in the URI are not shown properly formatted for escaping. Refer to Section 7.2 for the proper formatting. Additional detailed scenarios are available in the Appendix B.
Note: This example uses loose routing procedures.
INVITE sip:bob@biloxi.example.com;p=x
-------------->

Supported: histinfo

INVITE sip:bob@biloxi.example.com;p=x

History-Info: <sip:bob@biloxi.example.com;p=x>;index=1
History-Info: <sip:bob@biloxi.example.com;p=x>;index=1.1

INVITE sip:bob@192.0.2.3

History-Info: <sip:bob@biloxi.example.com;p=x>;index=1
History-Info: <sip:bob@biloxi.example.com;p=x>;index=1.1
History-Info: <sip:bob@192.0.2.3>;index=1.1;rc=1.1

200

History-Info: <sip:bob@biloxi.example.com;p=x>;index=1
History-Info: <sip:bob@biloxi.example.com;p=x>;index=1.1
History-Info: <sip:bob@192.0.2.3>;index=1.1;rc=1.1

Proxy Cancels INVITE

<------------------------>

200

History-Info: <sip:bob@biloxi.example.com;p=x>;index=1
History-Info: <sip:bob@biloxi.example.com;p=x>;index=1.1
History-Info: <sip:bob@192.0.2.3>;index=1.1;rc=1.1

ACK

-------------->

ACK

-------------->

--------------->
5. General User Agent Behavior

This section describes the processing specific to UAs for the History-Info header.

5.1. User Agent Client (UAC) Behavior

The UAC MUST include the "histinfo" option tag in the Supported header in any new or out-of-dialog request for which the UAC would like the History-Info header in the response. In addition, the UAC SHOULD add a History-Info header, using the Request-URI of the request as the hi-targeted-to-uri, in which case the index MUST be set to a value of 1 in the hi-entry. As a result, intermediaries and the UAS at least know the original Request-URI, and if the Request-URI was modified by a previous hop. In the case of a B2BUA implementation, a UAC MAY add the hi-entries received in the incoming request at the UAS to the subsequent outgoing request.

A UAC that does not want an hi-entry added due to privacy considerations MUST include a Privacy header with a priv-value(s) of "header" or "history". A UAC that wants to ensure that privacy not be applied to its identity MUST include a Privacy header with a priv-value of "none".

In the case where a UAC receives a 3xx response with a Contact header and sends a new request in response to it, the UAC MAY include in the outgoing request the previous hi-entry(s) received in the response. In this case, the reason header MUST be associated with the hi-targeted-to-uri in the previous (last) hi-entry, as described in Section 7.3.2. A new hi-entry MUST then be added for the URI from the Contact header (which becomes the new Request-URI). An index MUST be added to the hi-entry. The value for the index is determined following the rules for "Retargeting based upon a Response" as prescribed in Section 7.3.3. If the Contact header contained any of the header parameter fields defined in this specification, the UAC MUST include the header parameter field as a header parameter field associated with the current hi-entry as described in Section 7.3.4.

5.2. User Agent Server (UAS) Behavior

5.2.1. Processing of Requests with History-Info

Once the request terminates at the UAS, the UAS evaluates the History-Info header. The last hi-entry reflects the most recent target and MUST contain the Request-URI for the received request,
unless the previous entity that forwarded the request does not support the History-Info header. If the Request-URI of the incoming request does not match the last hi-entry (e.g., the last proxy does not support History-Info), the UAS MUST insert an hi-entry. The UAS MUST set the hi-targeted-to-uri based to the value of Request-URI in the incoming request. If privacy is required, a privacy header with a value of "history" MUST be added to the hi-entry. The UAS MUST include an hi-index attribute as described in Section 7.3.3. The UAS MUST NOT include a hi-target attribute, since the UAS has no way to know the mechanism by which the Request-URI was determined. The addition of the missing hi-entry ensures that the most complete information can be provided in the response and provides consistency in the information presented to applications. The information can also be useful for implementations with B2BUAs that include the History-Info, received in the incoming request, in the outgoing request.

Prior to any application usage of the information, the validity MUST be ascertained. If gaps are detected, this MUST NOT be treated as an error since gaps are possible if the request is forwarded through intermediate entities that do not support the History-Info header. The interpretation of the information in the History-Info header by a UAS in a request depends upon the specific applications supported by the UAS; an application might need to provide special handling in some cases where there are gaps. Application considerations and guidelines are provided in section 7.

5.2.2. Generation of Responses with History-Info

If the "histinfo" option tag is received in a request, the UAS MUST include any History-Info received in the request in the subsequent response. If privacy is required, entries MUST be anonymized as described in Section 7.3.1. The UAS MUST follow the rules for a redirect server per Section 5.3 in generating a 3xx response.

The processing of History-Info in responses follows the methodology described in Section 16.7 of [RFC3261], with the processing of History-Info headers adding an additional step, just before Step 9, "Forwarding the Response".

5.3. Redirect Server Behavior

A redirect server MUST include the History-Info headers received in the request in the 3XX response that it sends. A redirect server MUST add any new History-Info entries in cases of retargeting (both internal and to other SIP entities) in the same manner as prescribed for a proxy Section 6. In generating the Contact header in the 3xx response, the redirect server MUST add the appropriate target header.
field parameter to each Contact header as described in Section 7.3.4.

6. Proxy Behavior

This section describes the procedures for proxies and other SIP intermediaries for adding History-Info headers when requests are retargeted.

6.1. Adding the History-Info Header to Requests

This section describes the process of adding the History-Info header to requests for the following cases:

- Forwarding of initial request (see Section 6.1.1)
- Resending based on failure response (see Section 6.1.2)
- Resending based on redirection response (see Section 6.1.3)

Retargeting is an iterative process, i.e., a proxy may redirect "internally" more than one time. A typical example would be a proxy that retargets a request first to a different user (i.e., it maps to a different AOR), and then forwards to a registered contact bound to that new AOR. In these cases, a proxy MUST add multiple hi-entry fields. For example, a proxy that retargets bob@example.com to office@example.com and then to office@192.0.2.5 adds hi-entries for both office@example.com and office@192.0.2.5, in order to provide a logical description of the retargeting process internal to the proxy. A Reason MAY be associated with the hi-targeted-to-uri that has been retargeted as shown in the example in Appendix B.1.

6.1.1. Initial Request

Upon receipt of an initial request for a dialog, or a standalone request, a proxy forwarding the request MUST perform the following steps. Note that those steps below do not apply if the request is being re-sent as a result of failure (i.e., timeout, reception of an error response).

Step 1: Adding Entries on Behalf of Previous Hops

If an incoming request does not already have a History-Info header field (e.g., the UAC does not include any History-Info header and no proxies in between support History-Info), or if the Request-URI of the incoming request does not match the last hi-entry (e.g., the last proxy does not support History-Info), the proxy MUST insert an hi-entry. The proxy MUST set the hi-targeted-to-uri
based to the value of Request-URI in the incoming request, unless privacy is required. If privacy is required, the procedures of Section 7.3.1 MUST be used. The proxy MUST NOT include a hi-target attribute. The proxy MUST include an hi-index attribute with a value of "1", as described in Section 7.3.3.

Step 2: Generating New Entries for Each Outgoing Request

The proxy then proceeds to request forwarding as per 16.6/ [RFC3261]. The proxy MUST add a separate hi-entry in each separate outgoing request for each of the current (outgoing) targets in the target set. The proxy MUST set the hi-targeted-to-uri in those separate hi-entry(s) to the value of the Request-URI of the current (outgoing) request, unless privacy is required. If privacy is required, the procedures of Section 7.3.1 MUST be used. The proxy MUST include an hi-index for each of the separate hi-entry(s) as described in Section 7.3.3. The proxy MUST include the appropriate hi-target header field parameter for each of the separate entry(s) as described in Section 7.3.4, if applicable.

6.1.2. Re-sending based on failure response

When re-sending a request as a result of retargeting because of failure (i.e., either reception of error responses or a timeout which is considered to be an implicit 408 error response), the proxy MUST perform the following steps:

Step 1: Including the Entries from Error Responses & Timeouts

The proxy MUST build the History-Info header field(s) sent in the outgoing request using the aggregate information associated with the received error responses(s) and timeout(s) for all the branches that are generating failures, including the header entries in the order indicated by the indexing (see Section 7.3.3). If the received error response did not include any History-Info header fields, the proxy MUST use the same History-Info header fields that were sent in the outgoing request that failed to build the outgoing request.

Step 2: Tagging the Last Entries

The proxy then examines the last hi-entry of the History-Info that was just generated in Step 1 for each one of the branches that generated failures or timeouts and MUST add a Reason header for each one of those entries as per the procedures of Section 7.3.2.
Step 3: Generating New Entries for Each Outgoing Requests

Same as per Step 2 above for the normal forwarding case Section 6.1.1.

6.1.3. Re-sending based on redirection response

When re-sending a request as a result of retargeting because of redirection (i.e., receipt of a 3XX response), the following steps apply:

Step 1: Including Previous Entries

The proxy MUST include the History-Info header fields that were sent in the outgoing request that is being redirected.

Step 2: Tagging the Last Entry

The proxy then examines the last hi-entry of the History-Info that was just generated in Step 1 and MUST add a Reason header this entry as per the procedures of Section 7.3.2.

Step 3: Generating New Entries for Each Outgoing Requests

Same as per Step 2 for the normal forwarding case Section 6.1.1.

6.2. Sending History-Info in Responses

A proxy that receives a request with the "histinfo" option tag in the Supported header, MUST forward captured History-Info in subsequent, provisional, and final responses to the request sent by the ultimate UAS (see Section 5.2). Any hi-entry containing a Privacy header with a value of "history" MUST be anonymized prior to a proxy sending a response with that hi-entry.

The processing of History-Info in responses follows the methodology described in Section 16.7 of [RFC3261], with the processing of History-Info headers adding an additional step, just before Step 9, "Forwarding the Response".

7. The History-Info header field

7.1. Definition

History-Info is a header field as defined by [RFC3261]. It may appear in any initial request for a dialog, standalone request or responses associated with these requests. For example, History-Info
may appear in INVITE, REGISTER, MESSAGE, REFER, OPTIONS, SUBSCRIBE, and PUBLISH and any valid responses, plus NOTIFY requests that initiate a dialog.

The History-Info header carries the following information when the header is included in a request or response:

- **Targeted-to-URI (hi-targeted-to-uri):** A mandatory parameter for capturing the Request-URI for the specific request as it is forwarded.

- **Index (hi-index):** A mandatory parameter for History-Info reflecting the chronological order of the information, indexed to also reflect the forking and nesting of requests. The format for this parameter is a string of digits, separated by dots to indicate the number of forward hops and retargets. This results in a tree representation of the history of the request, with the lowest-level index reflecting a branch of the tree. By adding the new entries in order (i.e., following existing entries per the details in Section 6.1), including the index and securing the header, the ordering of the History-Info headers in the request is assured (SEC-req-2, see Appendix A.1). In addition, applications may extract a variety of metrics (total number of retargets, total number of retargets from a specific branch, etc.) based upon the index values.

- **Reason:** An optional parameter for History-Info, reflected in the History-Info header by including the Reason Header [RFC3326] escaped in the hi-targeted-to-uri. A reason is included for the hi-targeted-to-uri that was retargeted as opposed to the hi-targeted-to-uri to which it was retargeted.

- **Privacy:** An optional parameter for History-Info, reflected in the History-Info header field values by including the Privacy Header [RFC3323] escaped in the hi-targeted-to-uri or by adding the Privacy header to the request. The latter case indicates that the History-Info entries for the domain MUST be anonymized prior to forwarding, whereas the use of the Privacy header escaped in the hi-targeted-to-uri means that a specific hi-entry MUST be anonymized.

- **Target (hi-target):** An optional parameter for the History-Info and Contact headers. The hi-target is added for a hi-entry when it is first added in a History-Info header field, and only one value is permitted. Upon receipt of a request or response containing the History-Info header, a UA can determine the mechanism by which the target was determined. The following header field parameters are defined for this parameter:
"rc": The hi-targeted-to-URI is a contact for the Request-URI, in the incoming request, that is bound to an AOR in an abstract location service. The AOR-to-contact binding has been placed into the location service by a SIP Registrar that received a SIP REGISTER request. The "rc" header field parameter contains the index of the hi-entry associated with the URI in the incoming request.

"mp": The hi-targeted-to-URI represents a user other than the user associated with the Request-URI in the incoming requesting. This occurs when a request is to be statically or dynamically retargeted to another user. The value of the "mp" header field parameter is the index parameter for the hi-targeted-to-uri that was retargeted, thus identifying the "mapped from" target.

- Extension (hi-extension): A parameter to allow for future optional extensions. As per [RFC3261], any implementation not understanding an extension MUST ignore it.

The ABNF syntax for the History-Info header is defined as follows:

```
History-Info = "History-Info" HCOLON hi-entry *(COMMA hi-entry)
hi-entry = hi-targeted-to-uri *(SEMI hi-param)
hi-targeted-to-uri = name-addr
hi-param = hi-index / hi-target / hi-extension
index-val = 1*DIGIT *(
             
             
             )
hi-index = "index" EQUAL index-val
hi-target = rc-param / mp-param
rc-param = "rc" EQUAL index-val
mp-param = "mp" EQUAL index-val
hi-extension = generic-param
```

The ABNF definitions for "generic-param" and "name-addr" are from [RFC3261].

Note that since both the Reason and Privacy parameters are escaped in the hi-targeted-to-uri, these fields will not be available in the
case that the hi-targeted-to-uri is a Tel-URI [RFC3966].

7.2. Examples

The following provides some examples of the History-Info header. Note that the backslash and CRLF between the fields in the examples below are for readability purposes only.

History-Info: <sip:UserA@ims.example.com>;index=1;foo=bar

History-Info: <sip:UserA@ims.example.com?Reason=SIP%3B\cause%3D302>;index=1.1,
<sip:UserB@example.com?Privacy=history&Reason=SIP%3B\cause%3D486>;index=1.2;mp=1.1,
<sip:45432@192.168.0.3>;index=1.3;rc=1.2

7.3. Procedures

The following sections define procedures for processing of the parameters (and headers) associated with the History-Info header. These procedures may be applicable to processing entities such as Proxies, Redirect Servers or User Agents.

7.3.1. Privacy in the History-Info Header

The privacy requirements for this document are described in Appendix A.2.

As with other SIP headers described in [RFC3323], the History-Info header can inadvertently reveal information about the originator. A value of "header" in the Privacy header is used to indicate that such headers in the request be privacy protected when the request is forwarded by the intermediary. A value of "history" in the Privacy header indicates that the History-Info header should be privacy protected. If there is a Privacy header in the request with a priv-value of "header" or "history", then the initial hi-entry MUST be anonymized and the header removed when the request leaves a domain for which the SIP entity is responsible.

In addition, the History-Info header can reveal general routing and diverting information within an intermediary, which the intermediary may want to privacy protect. In this case, a value of "history" in the Privacy header is used to indicate which History-Info header entries added by a SIP entity are to be anonymized. The priv-value of "history" MUST be in the privacy header escaped in the hi-targeted-to-uri for each hi-entry, added by the SIP entity, as the request is retargeted within the domain for which the SIP entity is responsible. If a request is being retargeted to a URI associated
with a domain for which the SIP entity is not responsible, the processing entity MUST anonymize the hi-entries with a priv-value of "history" and MUST remove the Privacy header from the hi-entries prior to forwarding, unless the processing entity knows a priori that it can rely on a downstream processing entity to apply the requested privacy. The mechanism for the latter functionality is outside the scope of this specification.

Finally, the terminator of the request may not want to reveal the final reached target to the originator. In this case, the terminator uses the a value of "history" in the Privacy header in the last hi-entry in the response. The SIP entity that forwards the response MUST anonymize that hi-entry and remove the Privacy header.

7.3.2. Reason in the History-Info Header

For retargets that are the result of an explicit SIP response, a Reason MUST be associated with the hi-targeted-to-uri. If the SIP response does not include a Reason header (see [RFC3326]), the SIP Response Code that triggered the retargeting MUST be included as the Reason associated with the hi-targeted-to-uri that has been retargeted. If the response contains a Reason header for a protocol that is not SIP (e.g., Q.850), it MUST be captured as an additional Reason associated with the hi-targeted-to-uri that has been retargeted, along with the SIP Response Code. If the Reason header is a SIP reason, then it MUST be used as the Reason associated with the hi-targeted-to-uri rather than the SIP response code.

If a request has timed out (instead of being explicitly rejected), it MUST be treated as if a 408 "Request Terminated" error response code was received.

7.3.3. Indexing in the History-Info Header

In order to maintain ordering and accurately reflect the nesting and retargeting of the request, an index MUST be included along with the Targeted-to-URI being captured. Per the syntax in Section 7, the index consists of a dot-delimited series of digits (e.g., 1.1.2). Each dot reflects a hop or level of nesting; thus, the number of hops is determined by the total number of dots. Within each level, the integer reflects the number of peer entities to which the request has been routed. Thus, the indexing results in a logical tree representation for the history of the request.

The first index in a series of History-Info entries MUST be set to 1. In the case that a proxy adds an entry on behalf of the previous hop, the index MUST be set to 1. For each level of indexing, the index MUST start at 1. An increment of 1 MUST be used for advancing to a
new branch.

The basic rules for adding the index are summarized as follows:

1. **Basic Forwarding:** In the case of a request that is being forwarded, the index is determined by adding another sub-level of indexing since the depth/length of the branch is increasing. To accomplish this, the processing entity **MUST** read the value from the History-Info header in the received request and **MUST** add another level of indexing by appending the dot delimiter followed by an initial index for the new level of 1. For example, if the index in the last History-Info header field in the received request is 1.1, this proxy would initialize its index to 1.1.1 and forward the request.

2. **Retargeting within a processing entity - 1st instance:** For the first instance of retargeting within a processing entity, the calculation of the index follows that prescribed for basic forwarding.

3. **Retargeting within a processing entity - subsequent instance:** For each subsequent retargeting of a request by the same processing entity, another branch **MUST** be added. The index for each new branch **MUST** be calculated by incrementing the last/lowest digit at the current level, the index in the next request forwarded by this same processing entity, following the example above, would be 1.1.2.

4. **Retargeting based upon a Response:** In the case of retargeting due to a specific response (e.g., 302), the index **MUST** be calculated per rule 3. That is, the lowest/last digit of the index **MUST** be incremented (i.e., a new branch is created), with the increment of 1. For example, if the index in the History-Info header of the sent request is 1.2 and the response to the request is a 302, then the index in the History-Info header field for the new hi-targeted-to-URI would be 1.3.

5. **Forking requests:** If the request forwarding is done in multiple forks (sequentially or in parallel), the index **MUST** be captured for each forked request per the rules above, with each new request having a unique index. Each index **MUST** be sequentially assigned. For example, if the index in the last History-Info header field in the received request is 1.1, this processing entity would initialize its index to 1.1.1 for the first fork, 1.1.2 for the second, and so forth (see Figure 1 for an example). Note that for each individual fork, only the entry corresponding to that fork is included (e.g., the entry for fork 1.1.1 is not included in the request sent to fork 1.1.2, and vice-versa).
6. When a response is built and it represents the aggregate of responses to multiple forks (e.g., multiple forks that fail), the processing entity MUST build the subsequent response using the aggregated information associated with each of the forks for which there have been responses and MUST include the header entries in the order indicated by the indexing. For example, if a processing entity received failure responses for forks 1.1.1 and 1.1.2, it would return both the 1.1.1 and 1.1.2 entries to the entity that generated the hi-entry with index of 1. See Appendix B.1 for an example. Note that in the case of parallel forking where one fork is successful, only the forks for which responses have been received at the time the proxy sends the successful response are included in that response. Responses are processed as described in Section 16.7 of [RFC3261] with the aggregated History-Info entries processed similar to Step 7 "Aggregate Authentication Header Field Values".

7.3.4. Request Target in the History-Info Header

This specification defines two header field parameters, "rc" and "mp", indicating two non-inclusive mechanisms by which a new target for a request is determined. The specific parameter field to be included in the History-Info header is determined as the targets are added to the target set per the procedures of 16.5 or when the Contact header field in a 3xx response is populated. Both parameters contain an index whose value corresponds to the index in the hi-entry containing the Request-URI that was retargeted.

The specific parameter field to be included in the History-Info header is determined as follows:

- **"rc"**: If the hi-targeted-to-URI was determined based on a contact for the Request-URI being retargeted that is bound to an AOR in an abstract location service, then the "rc" header field parameter MUST be included in the hi-entry. The index of the "rc" parameter MUST be set to the hi-index containing the Request-URI being retargeted.

- **"mp"**: If the hi-targeted-to-URI was determined based on a mapping to a user other than the user associated with the Request-URI being retargeted, then the "mp" header field parameter MUST be included in the hi-entry. The index of the "mp" parameter MUST be set to the hi-index containing the Request-URI being retargeted.

Note that there are two scenarios by which the "mp" parameter can be derived.
The mapping was done by the receiving entity on its own authority, in which case the mp-value is the parent index of the hi-entry’s index.

The mapping was done due to receiving a 3xx response, in which case the mp-value is an earlier sibling of the hi-entry’s index, that of the downstream request which received the 3xx response.

8. Application Considerations

History-Info provides a very flexible building block that can be used by intermediaries and UAs for a variety of services. Prior to any application usage of the information the entries MUST be evaluated to determine gaps in indices. If gaps are detected, this MUST NOT be treated as an error since gaps are possible if the request is forwarded through intermediate entities that do not support the History-Info header. The interpretation of the information in the History-Info header depends upon the specific application; an application might need to provide special handling in some cases where there are gaps.

The following summarizes the categories of information that applications may use:

1. Complete history information – e.g., for debug or other operational and management aspects, optimization of determining targets to avoid retargeting to the same URI, etc. This information is relevant to proxies, UACs and UASs.

2. Entry with the index that matches the value of the last entry with a "rc" header parameter in the Request received by a UAS – i.e., the Request URI associated with the destination of the request was determined based on an AOR-to-contact binding in an abstract location service.

3. Entry with the index that matches the value of the last entry with a "mp" header parameter in the Request received by a UAS – i.e., the last Request URI that was mapped to reach the destination.

4. Entry with the index that matches the value of the first entry with a "rc" header parameter in the Request received by a UAS. Note, this would be the original AoR if all entries support the History-Info header and there is absence of a "mp" header parameter prior to the "rc" header parameter in the History-Info header. However, there is no guarantee that all entities will support History-Info, thus the first entry with an "rc" header...
parameter within the domain associated with the target URI at the destination is more likely to be useful.

5. Entry with the index that matches the value of the first entry with a "mp" header parameter in the Request received by a UAS. Note, this would be the original mapped URI if all entities supported the History-Info header. However, there is no guarantee that all entities will support History-Info, thus the first entry with an "mp" header parameter within the domain associated with the target URI at the destination is more likely to be useful.

In many cases, applications are most interested in the information within a particular domain(s), thus only a subset of the information is required.

Some applications may use multiple types of information. For example, an Automatic Call Distribution (ACD)/Call center application that utilizes the entry who index matches the index of the first History-Info entry with an hi-target value of "mp", may also display other agents, reflected by other History-Info entries prior to entries with hi-target values of "rc", to whom the call was targeted prior to its arrival at the current agent. This could allow the agent the ability to decide how they might forward or reroute the call if necessary (avoiding agents that were not previously available for whatever reason, etc.).

Since support for History-Info header is optional, a service MUST define default behavior for requests and responses not containing History-Info headers. For example, an entity may receive only partial History-Info entries or entries which are not tagged appropriately with an hi-target parameter. This may not impact some applications (e.g., debug), however, it could require some applications to make some default assumptions in this case. For example, in an ACD scenario, the application could select the oldest hi-entry with the domain associated with the ACD system and display that as the original called party. Depending upon how and where the request may have been retargeted, the complete list of agents to whom the call was targeted may not be available.

9. Security Considerations

The security requirements for this document are specified in Appendix A.1.

This document defines a header for SIP. The use of the Transport Layer Security (TLS) protocol [RFC5246] as a mechanism to ensure the
overall confidentiality of the History-Info headers (SEC-req-4) is strongly RECOMMENDED. This results in History-Info having at least the same level of security as other headers in SIP that are inserted by intermediaries. With TLS, History-Info headers are no less, nor no more, secure than other SIP headers, which generally have even more impact on the subsequent processing of SIP sessions than the History-Info header.

With the level of security provided by TLS (SEC-req-3), the information in the History-Info header can thus be evaluated to determine if information has been removed by evaluating the indices for gaps (SEC-req-1, SEC-req-2). It would be up to the application to define whether it can make use of the information in the case of missing entries.

Note that while using the SIPS scheme (as per [RFC5630]) protects History-Info from tampering by arbitrary parties outside the SIP message path, all the intermediaries on the path are trusted implicitly. A malicious intermediary could arbitrarily delete, rewrite, or modify History-Info. This specification does not attempt to prevent or detect attacks by malicious intermediaries.

10. IANA Considerations

This document requires several IANA registrations detailed in the following sections.

This document updates [RFC4244] but uses the same SIP header field name and option tag. The IANA registry needs to update the references to [RFC4244] with [RFCXXXX].

10.1. Registration of New SIP History-Info Header

This document defines a SIP header field name: History-Info and an option tag: histinfo. The following changes have been made to http://www.iana.org/assignments/sip-parameters The following row has been added to the header field section:

<table>
<thead>
<tr>
<th>Header Name</th>
<th>Compact Form</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>History-Info</td>
<td>none</td>
<td>[RFCXXXX]</td>
</tr>
</tbody>
</table>

The following has been added to the Options Tags section:
When used with the Supported header, this option tag indicates the UAC supports the History Information to be captured for requests and returned in subsequent responses. This tag is not used in a Proxy-Require or Require header field since support of History-Info is optional.

10.2. Registration of "history" for SIP Privacy Header

This document defines a priv-value for the SIP Privacy header:

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Registrant</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>history</td>
<td>Privacy requested for History-Info header(s)</td>
<td>Mary Barnes</td>
<td>[RFC XXXX]</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:mary.barnes@polycom.com">mary.barnes@polycom.com</a></td>
<td></td>
</tr>
</tbody>
</table>

10.3. Registration of Header Field Parameters

This specification defines the following new SIP header field parameters in the SIP Header Field parameter sub-registry in the SIP Parameter Registry, http://www.iana.org/assignments/sip-parameters:

<table>
<thead>
<tr>
<th>Header Field</th>
<th>Parameter Name</th>
<th>Predefined Values</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>History-Info</td>
<td>mp</td>
<td>No</td>
<td>[RFC xxxx]</td>
</tr>
<tr>
<td>History-Info</td>
<td>rc</td>
<td>No</td>
<td>[RFC xxxx]</td>
</tr>
<tr>
<td>Contact</td>
<td>mp</td>
<td>No</td>
<td>[RFC xxxx]</td>
</tr>
<tr>
<td>Contact</td>
<td>rc</td>
<td>No</td>
<td>[RFC xxxx]</td>
</tr>
</tbody>
</table>

11. Acknowledgements

Jonathan Rosenberg et al produced the document that provided additional use cases precipitating the requirement for the new header parameters to capture the method by which a Request URI is determined. The authors would like to acknowledge the constructive feedback provided by Ian Elz, Paul Kyzivat, John Elwell, Hadriel Kaplan and Dale Worley.

Mark Watson, Cullen Jennings and Jon Peterson provided significant input into the initial work that resulted in the development of [RFC4244]. The editor would like to acknowledge the constructive feedback provided by Robert Sparks, Paul Kyzivat, Scott Orton, John Elwell, Nir Chen, Palash Jain, Brian Stucker, Norma Ng, Anthony Brown, Jayshree Bharatia, Jonathan Rosenberg, Eric Burger, Martin Dolly, Roland Jesske, Takuya Sawada, Sebastien Prouvost, and Sebastien Garcin in the development of [RFC4244].

The editor would like to acknowledge the significant input from Rohan Mahy on some of the normative aspects of the ABNF for [RFC4244], particularly around the need for and format of the index and around the security aspects.

12. Changes from RFC 4244

This RFC replaces [RFC4244].

Deployment experience with [RFC4244] over the years has shown a number of issues, warranting an update:

- In order to make [RFC4244] work in "real life", one needs to make "assumptions" on how History-Info is used. For example, many implementations filter out many entries, and only leave specific entries corresponding, for example, to first and last redirection. Since vendors uses different rules, it causes significant interoperability issues.

- [RFC4244] is overly permissive and evasive about recording entries, causing interoperability issues.

- The examples in the call flows had errors, and confusing because they often assume "loose routing".
12.1. Backwards compatibility

This specification is backwards compatible since [RFC4244] allows for the addition of new optional parameters. This specification adds an
optional SIP header field parameter to the History-Info and Contact headers. Entities that have not implemented this specification MUST ignore these parameters, however, per [RFC4244] an entity MUST NOT remove this parameter from an hi-entry.

13. Changes since last Version

NOTE TO THE RFC-Editor: Please remove this section prior to publication as an RFC.

Changes from 01 to 02:

1. Editorial nits/clarifications. [Issues: 1, 6, 17, 18, 21-23, 25, 26, 30-33, 35-37, 39, 40]

2. Removing extraneous 4244 text - e.g., errors in flows, "stronger" security, "session" privacy. [Issues: 3, 5, 7, 11]

3. Updated definition of "retarget" to be all encompassing - i.e., also includes internal changes of target URI. Clarified text for "internal retargeting" in proxy section. [Issues: 2, 8, 9]

4. Clarified that the processing for Proxies is equally applicable to other SIP intermediaries. [Issue: 9].

5. Changed more SHOULDs to MUSTs. [Issue: 10]

6. Fixes to Application considerations section. [Issues: 12-15]

7. Changed language in the procedure for Indexing to normative language.

8. Clarifications for UAC processing:
   * MUST add hi-entry. [Issue: 28]
   * Clarify applicability to B2BUA. [Issue: 29]
   * Fixed text for indexing for UAC in case of 3xx.

9. Changed "hit" URI parameter to header parameters: [Issues: 4, 40]
   * Added index to all target header parameters. [Issues: 41]
   * Updated all the relevant sections documenting setting and use of new header parameters. [Issue: 40]
10. Updated/clarified privacy handling. [Issue: 16]

11. Updated Redirect Server section to allow adding History-Info headers. [Issue: 24]

12. Added text around restrictions for Tel-URIs - i.e., no privacy or reason. [Issues: 4, 12]

13. Updated text for forking - what goes in response. [Issues: 19, 20]

Changes from 00 to 01:

1. Moved examples (except first) in appendix to a new (informational) document.

2. Updated UAS and UAC sections to clarify and expand on the handling of the History-Info header.

3. Updated the Application considerations section:
   * Included more detail with regards to how applications can make use of the information, in particular based on the new tags.
   * Removed privacy consideration (2nd bullet) since privacy is now accomplished by anonymizing rather than removal of entries.

Changes from (individual) barnes-sipcore-4244bis-03 to (WG) ietf-sipcore-4244bis-00:

1. Added a new SIP/SIPS URI parameter to tag the URIs as they are added to the target list and those returned in the contact header in a 3xx response.

2. Updated description of "target" parameter to use the new URI parameter value in setting the value for the parameter.

3. Clarified privacy.

4. Changed handling at redirect server to include the use of the new URI parameter and to remove the functionality of adding the History-Info entries (basically reverting to core 4244 processing).

5. Additional text to clarify that a service such as voicemail can be done in multiple ways.
6. Editorial changes including removal of some vestiges of tagging all entries (including the "aor" tag).

Changes from barnes-sipcore-4244bis-02 to 03:
1. Fixed problem with indices in example in voicemail example.
2. Removed oc and rt from the Hi-target parameter.
3. Removed aor tag
4. Added index parameter to "mp"
5. Added use-cases and call-flows from target-uri into appendix.

Changes from barnes-sipcore-4244bis-01 to 02:
1. Added hi-aor parameter that gets marked on the "incoming" hi-entry.
2. Hi-target parameter defined to be either rc, oc, mp, rt, and now gets included when adding an entry.
3. Added section on backwards compatibility, as well as added the recognition and handling of requests that do not support this specification in the appropriate sections.
4. Updated redirect server/3xx handling to support the new parameters - i.e., the redirecting entity must add the new entry since the proxy does not have access to the information as to how the Contact was determined.
5. Added section on normative differences between this document and RFC 4244.
6. Restructuring of document to be more in line with current IETF practices.
7. Moved Requirements section into an Appendix.
8. Fixed ABNF to remove unintended ordering requirement on hi-index that was introduced in attempting to illustrate it was a mandatory parameter.

Changes from barnes-sipcore-4244bis-00 to 01:
1. Clarified "retarget" definition.
2. Removed privacy discussion from optionality section - just refer to privacy section.

3. Removed extraneous text from target-parameter (leftover from sip-4244bis). Changed the terminology from the "reason" to the "mechanism" to avoid ambiguity with parameter.

4. Various changes to clarify some of the text around privacy.

5. Reverted proxy response handling text to previous form - just changing the privacy aspects to anonymize, rather than remove.

6. Other editorial changes to condense and simplify.

7. Moved Privacy examples to Appendix.

8. Added forking to Basic call example.

Changes from barnes-sipcore-4244bis-00 to 01:

1. Clarified "retarget" definition.

2. Removed privacy discussion from optionality section - just refer to privacy section.

3. Removed extraneous text from target-parameter (leftover from sip-4244bis). Changed the terminology from the "reason" to the "mechanism" to avoid ambiguity with parameter.

4. Various changes to clarify some of the text around privacy.

5. Reverted proxy response handling text to previous form - just changing the privacy aspects to anonymize, rather than remove.

6. Other editorial changes to condense and simplify.

7. Moved Privacy examples to Appendix.

8. Added forking to Basic call example.

Changes from barnes-sip-4244bis-00 to barnes-sipcore-4244bis-00:

1. Added tags for each type of retargeting including proxy hops, etc. - i.e., a tag is defined for each specific mechanism by which the new Request-URI is determined. Note, this is extremely helpful in terms of backwards compatibility.
2. Fixed all the examples. Made sure loose routing was used in all of them.

3. Removed example where a proxy using strict routing is using History-Info for avoiding trying same route twice.

4. Remove redundant Redirect Server example.

5. Index is now mandated to start at "1" instead of recommended.

6. Updated 3xx behavior as the entity sending the 3XX response MUST add the hi-target attribute to the previous hi-entry to ensure that it is appropriately tagged (i.e., it’s the only one that knows how the contact in the 3xx was determined.)

7. Removed lots of ambiguity by making many "MAYs" into "SHOULDs" and some "SHOULDs" into "MUSTs".

8. Privacy is now recommended to be done by anonymizing entries as per RFC 3323 instead of by removing or omitting hi-entry(s).

9. Requirement for TLS is now same level as per RFC 3261.

10. Clarified behavior for "Privacy" (i.e., that Privacy is for Hi-entries, not headers).

11. Removed "OPTIONALITY" as specific requirements, since it’s rather superflous.

12. Other editorial changes to remove redundant text/sections.

Changes from RFC4244 to barnes-sip-4244bis-00:

1. Clarified that HI captures both retargeting as well as cases of just forwarding a request.

2. Added descriptions of the usage of the terms "retarget", "forward" and "redirect" to the terminology section.

3. Added additional examples for the functionality provided by HI for core SIP.

4. Added hi-target parameter values to HI header to ABNF and protocol description, as well as defining proxy, UAC and UAS behavior for the parameter.

5. Simplified example call flow in section 4.5. Moved previous call flow to appendix.
6. Fixed ABNF per RFC4244 errata "dot" -> "." and added new parameter.

14. References

14.1. Normative References


14.2. Informative References


[RFC3969] Camarillo, G., "The Internet Assigned Number Authority (IANA) Uniform Resource Identifier (URI) Parameter Registry for the Session Initiation Protocol (SIP)",
Appendix A. Request History Requirements

The following list constitutes a set of requirements for a "Request History" capability.

1. CAPABILITY-req: The "Request History" capability provides a capability to inform proxies and UAs involved in processing a request about the history/progress of that request. Although this is inherently provided when the retarget is in response to a SIP redirect, it is deemed useful for non-redirect retargeting scenarios, as well.

2. GENERATION-req: "Request History" information is generated when the request is retargeted.
   A. In some scenarios, it might be possible for more than one instance of retargeting to occur within the same Proxy. A proxy MUST also generate Request History information for the 'internal retargeting'.
   B. An entity (UA or proxy) retargeting in response to a redirect or REFER MUST include any Request History information from the redirect/REFER in the new request.

3. ISSUER-req: "Request History" information can be generated by a UA or proxy. It can be passed in both requests and responses.

4. CONTENT-req: The "Request History" information for each occurrence of retargeting shall include the following:
   A. The new URI or address to which the request is in the process of being retargeted,
   B. The URI or address from which the request was retargeted, and whether the retarget URI was an AOR
   C. The mechanism by which the new URI or address was determined,
   D. The reason for the Request-URI or address modification,
   E. Chronological ordering of the Request History information.
5. REQUEST-VALIDITY-req: Request History is applicable to requests not sent within an early or established dialog (e.g., INVITE, REGISTER, MESSAGE, and OPTIONS).

6. BACKWARDS-req: Request History information may be passed from the generating entity backwards towards the UAC. This is needed to enable services that inform the calling party about the dialog establishment attempts.

7. FORWARDS-req: Request History information may also be included by the generating entity in the request, if it is forwarded onwards.

A.1. Security Requirements

The Request History information is being inserted by a network element retargeting a Request, resulting in a slightly different problem than the basic SIP header problem, thus requiring specific consideration. It is recognized that these security requirements can be generalized to a basic requirement of being able to secure information that is inserted by proxies.

The potential security problems include the following:

1. A rogue application could insert a bogus Request History entry either by adding an additional entry as a result of retargeting or entering invalid information.

2. A rogue application could re-arrange the Request History information to change the nature of the end application or to mislead the receiver of the information.

3. A rogue application could delete some or all of the Request History information.

Thus, a security solution for "Request History" must meet the following requirements:

1. SEC-req-1: The entity receiving the Request History must be able to determine whether any of the previously added Request History content has been altered.

2. SEC-req-2: The ordering of the Request History information must be preserved at each instance of retargeting.

3. SEC-req-3: The entity receiving the information conveyed by the Request History must be able to authenticate the entity providing the request.
4. SEC-req-4: To ensure the confidentiality of the Request History information, only entities that process the request SHOULD have visibility to the information.

It should be noted that these security requirements apply to any entity making use of the Request History information.

A.2. Privacy Requirements

Since the Request-URI that is captured could inadvertently reveal information about the originator, there are general privacy requirements that MUST be met:

1. PRIV-req-1: The entity retargeting the Request must ensure that it maintains the network-provided privacy (as described in [RFC3323]) associated with the Request as it is retargeted.

2. PRIV-req-2: The entity receiving the Request History must maintain the privacy associated with the information. In addition, local policy at a proxy may identify privacy requirements associated with the Request-URI being captured in the Request History information.

3. PRIV-req-3: Request History information subject to privacy shall not be included in outgoing messages unless it is protected as described in [RFC3323].

Appendix B. Example call flows

The scenarios in this section provide sample use cases for the History-Info header for informational purposes only. They are not intended to be normative. A basic forking use case is included, along with two use cases illustrating the use of the privacy.

B.1. Sequentially Forking (History-Info in Response)

This scenario highlights an example where the History-Info in the response is useful to an application or user that originated the request.

Alice sends a call to Bob via sip:example.com. The proxy sip:example.com sequentially tries Bob on a SIP UA that has bound a contact with the sip:bob@example.com AOR, and then several alternate addresses (Office and Home) unsuccessfully before sending a response to Alice. The hi-entry containing the initial contact is the entry just prior to the first entry tagged with an hi-target value of "rc". In this example, the Office and Home are not the same AOR as
sip:bob@example.com, but rather different AORs that have been configured as alternate addresses for Bob in the proxy. In other words, Office and Bob are not bound through SIP Registration with Bob’s AOR. This type of arrangement is common for example when a "routing" rule to a PSTN number is manually configured in a Proxy. These hi-entries are identified by the index contained in the hi-target "mp" parameter in the hi-entries.

This scenario illustrates that by providing the History-Info to Alice, the end-user or an application at Alice could make a decision on how best to attempt finding Bob without sending multiple requests to the same destination. Upon receipt of the response containing the History-Info entries, the Request URIs for the History-Info entries tagged with "mp" are extracted. Those Request-URIs can be compared to other URIs (if any) that might be attempted in order to establish the session with Bob. Thus, avoiding another INVITE to Bob’s home phone. Without this mechanism, Alice might well attempt to reach Bob at his office phone, which would then re-target the request to Bob’s home phone. When that attempt failed, then Alice might attempt to reach Bob directly at his home phone, unknowingly for a third time.
Message Details

F1 INVITE alice -> example.com

INVITE sip:alice@example.com SIP/2.0
Via: SIP/2.0/TCP 192.0.2.3:5060
From: Alice <sip:alice@example.com>
To: Bob <sip:bob@example.com>
Supported: histinfo
Call-Id: 12345600@example.com
CSeq: 1 INVITE
History-Info: <sip:bob@example.com>;index=1
Contact: Alice <sip:alice@192.0.2.3>
Content-Type: application/sdp
Content-Length: <appropriate value>
<!-- SDP Not Shown -->

F2 INVITE  example.com -> Bob

INVITE sip:bob@192.0.2.4 SIP/2.0
Via: SIP/2.0/TCP proxy.example.com:5060
Via: SIP/2.0/TCP 192.0.2.3:5060
From: Alice <sip:alice@example.com>
To: Bob <sip:bob@example.com>
Supported: histinfo
Call-Id: 12345600@example.com
CSeq: 1 INVITE
Record-Route: <sip:proxy.example.com;lr>
History-Info: <sip:bob@example.com>;index=1
History-Info: <sip:bob@192.0.2.4>;index=1.1;rc=1
Contact: Alice <sip:alice@192.0.2.3>
Content-Type: application/sdp
Content-Length: <appropriate value>
<!-- SDP Not Shown -->

F3 100 Trying example.com -> alice

SIP/2.0 100 Trying
Via: SIP/2.0/TCP 192.0.2.3:5060
From: Alice <sip:alice@example.com>
To: Bob <sip:bob@example.com>
Call-Id: 12345600@example.com
CSeq: 1 INVITE
Content-Length: 0

F4 302 Moved Temporarily Bob -> example.com

SIP/2.0 302 Moved Temporarily
Via: SIP/2.0/TCP proxy.example.com:5060
Via: SIP/2.0/TCP 192.0.2.3:5060
From: Alice <sip:alice@example.com>
To: Bob <sip:bob@example.com>;tag=3
Call-Id: 12345600@example.com
CSeq: 1 INVITE
Record-Route: <sip:proxy.example.com;lr>
History-Info: <sip:bob@example.com>;index=1
History-Info: <sip:bob@192.0.2.4>;index=1.1;rc=1
Contact: <sip:office@example.com>;mp=1
Content-Length: 0
F5 ACK 192.0.2.4 -> Bob

ACK sip:home@example.com SIP/2.0
Via: SIP/2.0/TCP proxy.example.com:5060
From: Alice <sip:alice@example.com>
To: Bob <sip:bob@example.com>
Call-ID: 12345600@example.com
CSeq: 1 ACK
Content-Length: 0

F6 INVITE example.com -> office

INVITE sip:office@192.0.2.3.com SIP/2.0
Via: SIP/2.0/TCP proxy.example.com:5060;branch=2
Via: SIP/2.0/TCP 192.0.2.3:5060
From: Alice <sip:alice@example.com>
To: Bob <sip:bob@example.com>
Supported: histinfo
Call-ID: 12345600@example.com
Record-Route: <sip:proxy.example.com;lr>
History-Info: <sip:bob@example.com>;index=1
History-Info: <sip:bob@192.0.2.4?Reason=SIP;cause=302>;index=1.1;rc=1
History-Info: <sip:office@example.com>;index=1.2;mp=1
History-Info: <sip:office@192.0.2.5>;index=1.2.1
CSeq: 1 INVITE
Contact: Alice <sip:alice@192.0.2.3>
Content-Type: application/sdp
Content-Length: <appropriate value>
<!-- SDP Not Shown -->
F7 180 Ringing office -> example.com

SIP/2.0 180 Ringing
Via: SIP/2.0/TCP proxy.example.com:5060;branch=2
Via: SIP/2.0/TCP 192.0.2.3:5060
From: Alice <sip:alice@example.com>
To: Bob <sip:bob@example.com>;tag=5
Supported: histinfo
Call-ID: 12345600@example.com
Record-Route: <sip:proxy.example.com;lr>
History-Info: <sip:bob@example.com>;index=1
History-Info: <sip:bob@192.0.2.4?Reason=SIP;cause=302>;\index=1.1;rc=1
History-Info: <sip:office@example.com>;index=1.2;mp=1
History-Info: <sip:office@192.0.2.5>;index=1.2.1
CSeq: 1 INVITE
Content-Length: 0

F8 180 Ringing example.com -> alice

SIP/2.0 180 Ringing
Via: SIP/2.0/TCP example.com:5060
From: Alice <sip:alice@example.com>
To: Bob <sip:bob@example.com>
Supported: histinfo
Call-ID: 12345600@example.com
History-Info: <sip:bob@example.com>;index=1
History-Info: <sip:bob@192.0.2.4?Reason=SIP;cause=302>;\index=1.1;rc=1
History-Info: <sip:office@example.com>;index=1.2;mp=1
History-Info: <sip:office@192.0.2.5>;index=1.2.1
CSeq: 1 INVITE
Content-Length: 0
F9 INVITE example.com -> home

INVITE sip:home@192.0.2.6 SIP/2.0
Via: SIP/2.0/TCP proxy.example.com:5060;branch=3
Via: SIP/2.0/TCP 192.0.2.3:5060
From: Alice <sip:alice@example.com>
To: Bob <sip:bob@example.com>
Supported: histinfo
Call-Id: 12345600@example.com
Record-Route: <sip:proxy.example.com;lr>
History-Info: <sip:bob@192.0.2.4?Reason=SIP;cause=302>;\  
index=1.1;rc=1
History-Info: <sip:office@example.com>;index=1.2;mp=1
History-Info: <sip:office@192.0.2.5?Reason=SIP;cause=480>;\  
index=1.2.1>;index=1.2.1
History-Info: <sip:home@example.com>;index=1.3;mp=1
History-Info: <sip:home@192.0.2.6>;index=1.3.1
CSeq: 1 INVITE
Contact: Alice <sip:alice@192.0.2.3>
Content-Type: application/sdp
Content-Length: <appropriate value>
<!-- SDP Not Shown -->

F10 100 Trying home -> example.com

SIP/2.0 100 Trying
Via: SIP/2.0/TCP proxy.example.com:5060;branch=3
Via: SIP/2.0/TCP 192.0.2.3:5060
From: Alice <sip:alice@example.com>
To: Bob <sip:bob@example.com>
Call-Id: 12345600@example.com
CSeq: 1 INVITE
Content-Length: 0
F11 486 Busy Here home -> example.com

SIP/2.0  486 Busy Here
Via: SIP/2.0/TCP proxy.example.com:5060;branch=3
Via: SIP/2.0/TCP 192.0.2.3:5060
From: Alice <sip:alice@example.com>
To: Bob <sip:bob@example.com>
Call-Id: 12345600@example.com
Record-Route: <sip:proxy.example.com;lr>
History-Info: <sip:bob@example.com>;index=1
History-Info: <sip:bob@192.0.2.4?Reason=SIP;cause=302>;
  index=1.1;rc=1
History-Info: <sip:office@example.com>;index=1.2;mp=1
History-Info: <sip:office@192.0.2.5?Reason=SIP;cause=480>;
  index=1.2.1;index=1.2.1
History-Info: <sip:home@example.com>;index=1.3;mp=1
History-Info: <sip:home@192.0.2.6>;index=1.3.1
CSeq: 1 INVITE
Content-Length: 0

F12 486 Busy Here example.com -> alice

SIP/2.0  486 Busy Here
Via: SIP/2.0/TCP 192.0.2.3:5060
From: Alice <sip:alice@example.com>
To: Bob <sip:bob@example.com>
Call-Id: 12345600@example.com
History-Info: <sip:bob@example.com>;index=1
History-Info: <sip:bob@192.0.2.4?Reason=SIP;cause=302>;
  index=1.1;rc=1
History-Info: <sip:office@example.com>;index=1.2;mp=1
History-Info: <sip:office@192.0.2.5?Reason=SIP;cause=480>;
  index=1.2.1;index=1.2.1
History-Info: <sip:home@example.com>;index=1.3;mp=1
History-Info: <sip:home@192.0.2.6>;index=1.3.1
CSeq: 1 INVITE
Content-Length: 0
**B.2. History-Info with Privacy Header**

This example provides a basic call scenario without forking. Alice has indicated that she wants Privacy associated with her History-Info entry and sip:biloxi.example.com adds Privacy headers indicating that the History-Info header information is anonymized outside the biloxi.example.com domain. Note, that if the atlanta.example.com proxy had added privacy headers to all its hi-entries, then all the hi-entries in the response would be anonymous.
Alice   atlanta.example.com  biloxi.example.com   Bob

INVITE sip:bob@biloxi.example.com;p=x

--------------->

Supported: histinfo
Privacy: History
History-Info: <sip:bob@biloxi.example.com;p=x>;index=1

INVITE sip:bob@biloxi.example.com;p=x

--------------->

History-Info: <sip:anonymous@anonymous.invalid>;index=1
History-Info: <sip:anonymous@anonymous.invalid>;index=1

INVITE sip:192.0.2.3

--------------->

History-Info: <sip:anonymous@anonymous.invalid>;index=1
History-Info: <sip:anonymous@anonymous.invalid>;index=1

200

--------------->

History-Info: <sip:anonymous@anonymous.invalid>;index=1
History-Info: <sip:anonymous@anonymous.invalid>;index=1

--------

200

--------------->

History-Info: <sip:anonymous@anonymous.invalid>;index=1
History-Info: <sip:anonymous@anonymous.invalid>;index=1

-------

200

--------------->

History-Info: <sip:anonymous@anonymous.invalid>;index=1
History-Info: <sip:anonymous@anonymous.invalid>;index=1

ACK

--------------->

ACK

--------------->

ACK

Figure 2: Example with Privacy Header
B.3. Privacy Header for a Specific History-Info Entry

This example provides a basic call scenario similar to Appendix B.2, however, due to local policy at sip:biloxi.example.com, only the final hi-entry in the History-Info, which is Bob’s local URI, contains a priv-value of "history", thus providing Alice with some information about the history of the request, but anonymizing Bob’s local URI.
Alice atlanta.example.com biloxi.example.com Bob

INVITE sip:bob@biloxi.example.com;p=x

--------------->

Supported: histinfo

INVITE sip:bob@biloxi.example.com;p=x

-------------->

History-Info: <sip:bob@biloxi.example.com;p=x>;index=1
History-Info: <sip:bob@biloxi.example.com;p=x>;index=1.1

-------------->

INVITE sip:bob@192.0.2.3

-------------->

History-Info: <sip:bob@biloxi.example.com;p=x>;index=1
History-Info: <sip:bob@biloxi.example.com;p=x>;index=1.1
History-Info: <sip:bob@192.0.2.3>;index=1.1;rc=1.1

-------------->

200

History-Info: <sip:bob@biloxi.example.com;p=x>;index=1
History-Info: <sip:bob@biloxi.example.com;p=x>;index=1.1
History-Info: <sip:anonymous@anonymous.invalid>;index=1.1;rc=1.1

-------------->

200

History-Info: <sip:bob@biloxi.example.com;p=x>;index=1
History-Info: <sip:bob@biloxi.example.com;p=x>;index=1.1
History-Info: <sip:anonymous@anonymous.invalid>;index=1.1;rc=1.1

-------------->

ACK

Figure 3: Example with Privacy Header for Specific URI
Authors’ Addresses

Mary Barnes
Polycom
TX
US

Email: mary.ietf.barnes@gmail.com

Francois Audet
Skype

Email: francois.audet@skype.net

Shida Schubert
NTT

Email: shida@agnada.com

Hans Erik van Elburg
Detecon International GmbH
Oberkasseler str. 2
Bonn,
Germany

Email: ietf.hanserik@gmail.com

Christer Holmberg
Ericsson
Hirsalantie 11, Jorvas
Finland

Email: christer.holmberg@ericsson.com
Example call flows using Session Initiation Protocol (SIP) security mechanisms
draft-ietf-sipcore-sec-flows-09

Abstract

This document shows example call flows demonstrating the use of Transport Layer Security (TLS), and Secure/Multipurpose Internet Mail Extensions (S/MIME) in Session Initiation Protocol (SIP). It also provides information that helps implementers build interoperable SIP software. To help facilitate interoperability testing, it includes certificates used in the example call flows and processes to create certificates for testing.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on August 13, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents
Table of Contents

1. Introduction .................................................. 3
2. Certificates ..................................................... 4
   2.1. CA Certificates .............................................. 4
   2.2. Host Certificates ........................................... 8
   2.3. User Certificates ........................................... 10
3. Callflow with Message Over TLS ................................. 12
   3.1. TLS with Server Authentication ............................. 12
   3.2. MESSAGE Transaction Over TLS ............................... 13
4. Callflow with S/MIME-secured Message ......................... 15
   4.1. MESSAGE Request with Signed Body .......................... 15
   4.2. MESSAGE Request with Encrypted Body ...................... 20
   4.3. MESSAGE Request with Encrypted and Signed Body .......... 22
5. Observed Interoperability Issues .............................. 29
6. Additional Test Scenarios ...................................... 31
7. IANA Considerations ............................................ 34
8. Acknowledgments ................................................ 35
9. Security Considerations ........................................ 36
10. Changelog ....................................................... 37
11. References ...................................................... 40
   11.1. Normative References ...................................... 40
   11.2. Informative References .................................... 41
Appendix A. Making Test Certificates ............................. 43
   A.1. makeCA script .............................................. 44
   A.2. makeCert script ............................................ 48
Appendix B. Certificates for Testing ............................. 51
   B.1. Certificates Using EKU ..................................... 51
   B.2. Certificates NOT Using EKU ................................ 58
   B.3. Certificate Chaining with a Non-Root CA .................. 66
Appendix C. Message Dumps ........................................... 73
Authors’ Addresses .................................................. 76
1. Introduction

This document is informational and is not normative on any aspect of SIP.

SIP with TLS ([RFC5246]) implementations are becoming very common. Several implementations of the S/MIME ([RFC5751]) portion of SIP ([RFC3261]) are also becoming available. After several interoperability events, it is clear that it is difficult to write these systems without any test vectors or examples of "known good" messages to test against. Furthermore, testing at the events is often hindered due to the lack of a commonly trusted certificate authority to sign the certificates used in the events. This document addresses both of these issues by providing messages that give detailed examples that implementers can use for comparison and that can also be used for testing. In addition, this document provides a common certificate and private key that can be used to set up a mock Certificate Authority (CA) that can be used during the SIP interoperability events. Certificate requests from the users will be signed by the private key of the mock CA. The document also provides some hints and clarifications for implementers.

A simple SIP call flow using SIPS URIs and TLS is shown in Section 3. The certificates for the hosts used are shown in Section 2.2, and the CA certificates used to sign these are shown in Section 2.1.

The text from Section 4.1 through Section 4.3 shows some simple SIP call flows using S/MIME to sign and encrypt the body of the message. The user certificates used in these examples are shown in Section 2.3. These host certificates are signed with the same mock CA private key.

Section 5 presents a partial list of items that implementers should consider in order to implement systems that will interoperate.

Scripts and instructions to make certificates that can be used for interoperability testing are presented in Appendix A, along with methods for converting these to various formats. The certificates used while creating the examples and test messages in this document are made available in Appendix B.

Binary copies of various messages in this document that can be used for testing appear in Appendix C.
2. Certificates

2.1. CA Certificates

The certificate used by the CA to sign the other certificates is shown below. This is a X509v3 certificate. Note that the X.509v3 Basic Constraints in the certificate allows it to be used as a CA, certificate authority. This certificate is not used directly in the TLS call flow; it is used only to verify user and host certificates.

Version: 3 (0x2)
Serial Number:
96:a3:84:17:4e:ef:8a:4c
Signature Algorithm: sha1WithRSAEncryption
Issuer: C=US, ST=California, L=San Jose, O=sipit, OU=Sipit Test Certificate Authority
Validity
Not Before: Jan 27 18:36:05 2011 GMT
Not After : Jan 3 18:36:05 2111 GMT
Subject: C=US, ST=California, L=San Jose, O=sipit, OU=Sipit Test Certificate Authority
Subject Public Key Info:
Public Key Algorithm: rsaEncryption
RSA Public Key: (2048 bit)
Modulus (2048 bit):
f8:34:41:70:d9:c0:03:91:6a:ba:d1:11:8f:ac:12:
5b:c2:de:0b:26:65:d0:91:c7:70:4b:c7:0a:4a:bf:
2c:78:ec:a5:0f:be:9c:10:f8:c0:0b:0d:73:99:9e:
29:c3
Exponent: 65537 (0x10001)
X509v3 extensions:
X509v3 Authority Key Identifier:
X509v3 Basic Constraints:
CA:TRUE

Signature Algorithm: sha1WithRSAEncryption

The certificate content shown above and throughout this document was rendered by the OpenSSL "x509" tool. These dumps are included only as informative examples. Output may vary among future revisions of the tool. At the time of this document’s publication, there were some irregularities in the presentation of Distinguished Names (DN). In particular, note that in the "Issuer" and "Subject" fields, it appears the intent is to present DNS in Lightweight Directory Access Protocol (LDAP) format. If this was intended, the spaces should have been omitted after the delimiting commas, and the elements should have been presented in order of most-specific to least-specific. Please refer to Appendix A of [RFC4514]. Using the "Issuer" DN from above as an example and following guidelines in [RFC4514], it should have instead appeared as:

Issuer: OU=Sipit Test Certificate Authority,O=sipit,L=San Jose,ST=California,C=US

The ASN.1 parse of the CA certificate is shown below.

0:l= 949 cons: SEQUENCE
4:l= 669 cons: SEQUENCE
8:l= 3 cons: cont [ 0 ]
10:l= 1 prim: INTEGER :02
13:l= 9 prim: INTEGER :96A384174EEF8A4C
24:l= 13 cons: SEQUENCE
26:l= 9 prim: OBJECT :sha1WithRSAEncryption
37:l= 0 prim: NULL
39:l= 112 cons: SEQUENCE
41:l= 11 cons: SET
43:l= 9 cons: SEQUENCE
45:l= 3 prim: OBJECT :countryName
50:l= 2 prim: PRINTABLESTRING :US
54:l= 19 cons: SET
56:l= 17 cons: SEQUENCE
58:l= 3 prim: OBJECT :stateOrProvinceName
63:l= 10 prim: UTF8STRING
   43 61 6c 69 66 6f 72 6e-69 61                     California
75:l= 17 cons: SET
77:l= 15 cons: SEQUENCE
79:l= 3 prim: OBJECT :localityName
84:l= 8 prim: UTF8STRING
   53 61 6e 20 4a 6f 73 65-                          San Jose
94:l= 14 cons: SET
96:l= 12 cons: SEQUENCE
98:l= 3 prim: OBJECT :organizationName
103:l= 5 prim: UTF8STRING
   73 69 70 69 74 sipit
110:l= 41 cons: SET
112:l= 39 cons: SEQUENCE
114:l= 3 prim: OBJECT :organizationalUnitName
119:l= 32 prim: UTF8STRING
   53 69 70 69 74 20 54 65-73 74 20 43 65 72 74 69-   Sipit Test Certi
   66 69 63 61 74 65 20 41-75 74 79 ficate Authority
153:l= 32 cons: SEQUENCE
155:l= 13 prim: UTCTIME :110127183605Z
170:l= 15 prim: GENERALIZEDTIME :21110103183605Z
187:l= 112 cons: SEQUENCE
189:l= 11 cons: SET
191:l= 9 cons: SEQUENCE
193:l= 3 prim: OBJECT :countryName
198:l= 2 prim: PRINTABLESTRING :US
202:l= 19 cons: SET
204:l= 17 cons: SEQUENCE
206:l= 3 prim: OBJECT :stateOrProvinceName
211:l= 10 prim: UTF8STRING
   43 61 6c 69 66 6f 72 6e-69 61                     California
223:l= 17 cons: SET
225:l= 15 cons: SEQUENCE
227:l= 3 prim: OBJECT :localityName
232:l= 8 prim: UTF8STRING
   53 61 6e 20 4a 6f 73 65-                          San Jose
242:l= 14 cons: SET
244:l= 12 cons: SEQUENCE
246:l= 3 prim: OBJECT :organizationName
251: l= 5 prim: UTF8STRING
   73 69 70 69 74 sipit
258: l= 41 cons: SET
260: l= 39 cons: SEQUENCE
262: l= 3 prim: OBJECT :organizationalUnitName
267: l= 32 prim: UTF8STRING
   53 69 70 69 74 20 54 65-73 74 20 43 65 6f 72 69 74 69 6e Test Certi
86 69 63 61 74 65 20 41 75 74 68 6f 72 69 74 79 ficate Authority
301: l= 290 cons: SEQUENCE
305: l= 13 cons: SEQUENCE
307: l= 9 prim: OBJECT :rsaEncryption
318: l= 0 prim: NULL
320: l= 271 prim: BIT STRING
   00 30 82 01 0a 02 82 01-01 00 ab 1f 91 61 f1 1c .0...........a..
   c5 cd a6 7b 16 9b b7 14-79 e4 30 9e 98 d0 ec 07 ...{[.]y.0.....
   b7 bd 77 d7 d1 f5 5b 2c-e2 ee e6 b1 b0 f0 85 fa ...\........
   a5 bc cb cc cf 69 2c 4f-fc 50 ef 9d 31 2b c0 59 ....i,0.p...+Y
   ea fb 64 6f 1f 55 a7 3d-fd 70 d2 56 db 14 99 17 ...do.U.=.p.V....
   9f 70 ac 26 f8 34 41 70-d9 c0 03 91 6a ba d1 11 .p.&.4Ap.......
   8f ac 12 31 de b9 19 70-8d 5d a7 7d 8b 19 cc 40 .....1..p.]}...@
   3f ae ff de 1f db 94 b3-46 77 6c ae ae ff 3e d6 ?.........FwL...
   84 5b c2 de 0b 26 65 d0-91 c7 70 4b c7 0a 4a bf [.][e...p.K....
   c7 97 0d dd ba 58 47 cb-e0 2b 23 76 87 65 c5 55 ....XG.+s#v.e.U
   34 10 ab 27 1f 1c f8 30-3d b0 9b ca a2 81 72 4c 4c."...0==...rL
   bd 60 fe f7 21 fe 0b db-0b db e9 5b 01 36 d4 28 `...!111bc[.6(.
   15 6b 79 eb d0 91 1b 21-59 b8 0e aa bf d5 b1 6c .ky...!Y.....l
   70 37 a3 3f a5 7d 0e 95-46 f6 f6 58 67 83 75 42 p7.?2.F.Xg.ub
   37 18 0b a4 41 39 b2 2f-6c 80 2c 78 ec a5 0f be 7...A9./,x....
   9c 10 f8 c0 0b 0d 73 99-9e 0d d7 97 50 cb cc 45 ......s...P.f.E
   34 23 49 41 85 22 24 ad-29 c3 02 03 01 00 01 4#IA."$....)

595: l= 80 cons: cont [ 3 ]
597: l= 78 cons: SEQUENCE
599: l= 29 cons: SEQUENCE
601: l= 3 prim: OBJECT :X509v3 Subject Key Identifier
606: l= 22 prim: OCTET STRING
   04 14 95 45 7e 5f 2b ea-65 98 12 91 04 f3 63 c7 ...E^+.e......c.
   68 9a 58 16 77 27 h.x.w'
630: l= 31 cons: SEQUENCE
632: l= 3 prim: OBJECT :X509v3 Authority Key Identifier
637: l= 24 prim: OCTET STRING
   30 16 80 14 95 45 7e 5f-2b ea 65 98 12 91 04 f3 0....E^+.e.....
   63 c7 68 9a 58 16 77 27-c.h.x.w'
663: l= 12 cons: SEQUENCE
665: l= 3 prim: OBJECT :X509v3 Basic Constraints
670: l= 5 prim: OCTET STRING
   30 03 01 01 ff 0....
677: l= 13 cons: SEQUENCE
679: l= 9 prim: OBJECT :sha1WithRSAEncryption
2.2. Host Certificates

The certificate for the host example.com is shown below. Note that the Subject Alternative Name is set to example.com and is a DNS type. The certificates for the other hosts are shown in Appendix B.

Version: 3 (0x2)
Serial Number:
96:a3:84:17:4e:ef:8a:4f
Signature Algorithm: sha1WithRSAEncryption
Issuer: C=US, ST=California, L=San Jose, O=sipit, OU=Sipit Test Certificate Authority
Validity
Not Before: Feb 7 19:32:17 2011 GMT
Not After : Jan 14 19:32:17 2111 GMT
Subject: C=US, ST=California, L=San Jose, O=sipit, CN=example.com
Subject Public Key Info:
Public Key Algorithm: rsaEncryption
RSA Public Key: (2048 bit)
Modulus (2048 bit):
00:dd:74:06:02:10:c2:e7:04:1f:bc:8c:b6:24:e7:
The example host certificate above, as well as all the others presented in this document, are signed directly by a root CA. These certificate chains have a length equal to two: the root CA and the host certificate. Non-root CAs exist and may also sign certificates.
The certificate chains presented by hosts with certificates signed by non-root CAs will have a length greater than two. For more details on how certificate chains are validated, see Sections 6.1 and 6.2 of [RFC5280].

2.3. User Certificates

User certificates are used by many applications to establish user identity. The user certificate for fluffy@example.com is shown below. Note that the Subject Alternative Name has a list of names with different URL types such as a sip, im, or pres URL. This is necessary for interoperating with a Common Profile for Instant Messaging (CPIM) gateway. In this example, example.com is the domain for fluffy. The message could be coming from any host in *.example.com, and the AOR in the user certificate would still be the same. The others are shown in Appendix B.1. These certificates make use of the Extended Key Usage (EKU) extension discussed in [RFC5924]. Note that the X509v3 Extended Key Usage attribute refers to the SIP OID introduced in [RFC5924], which is 1.3.6.1.5.5.7.3.20

Version: 3 (0x2)
Serial Number: 96:a3:84:17:4e:ef:8a:4d
Signature Algorithm: sha1WithRSAEncryption
Issuer: C=US, ST=California, L=San Jose, O=sipit, OU=Sipit Test Certificate Authority
Validity
Not Before: Feb 7 19:32:17 2011 GMT
Not After : Jan 14 19:32:17 2111 GMT
Subject: C=US, ST=California, L=San Jose, O=sipit, CN=fluffy
Subject Public Key Info:
Public Key Algorithm: rsaEncryption
RSA Public Key: (2048 bit)
Modulus (2048 bit):
00:a3:2c:59:0c:e9:bc:e4:ec:d3:9e:fb:99:02:ec:
05:61:0b:0a:ca:ca:ec:51:ec:53:6e:3d:2b:00:80:
fe:35:1b:06:0a:61:13:88:08:44:03:cc:fd:2b:0e:
X509v3 extensions:

X509v3 Subject Alternative Name:
URI:sip:fluffy@example.com, URI:im:fluffy@example.com,
URI:pres:fluffy@example.com

X509v3 Basic Constraints:
CA:FALSE

X509v3 Subject Key Identifier:

X509v3 Authority Key Identifier:

X509v3 Key Usage:

Signature Algorithm: sha1WithRSAEncryption

E-mail Protection, 1.3.6.1.5.5.7.3.20

Versions of these certificates that do not make use of EKU are also included in Appendix B.2
3. Callflow with Message Over TLS

3.1. TLS with Server Authentication

The flow below shows the edited SSLEDump output of the host example.com forming a TLS [RFC5246] connection to example.net. In this example mutual authentication is not used. Note that the client proposed three protocol suites including TLS_RSA_WITH_AES_128_CBC_SHA defined in [RFC5246]. The certificate returned by the server contains a Subject Alternative Name that is set to example.net. A detailed discussion of TLS can be found in SSL and TLS [EKR-TLS]. For more details on the SSLEDump tool, see the SSLEDump Manual [ssledump-manpage].

This example does not use the Server Extended Hello (see [RFC5246]).

New TCP connection #1: example.com(50738) <-> example.net(5061)

1 1 0.0004 (0.0004) C>SV3.1(101) Handshake
ClientHello
  Version 3.1
  random[32]=
    4c 09 5b a7 66 77 eb 43 52 30 dd 98 4d 09 23 d3
    ff 81 74 ab 04 69 bb 79 8c dc 59 cd c2 1f b7 ec
  cipher suites
    TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA
    TLS_ECDH_RSA_WITH_AES_256_CBC_SHA
    TLS_DHE_RSA_WITH_AES_256_SHA
    TLS_RSA_WITH_AES_256_CBC_SHA
    TLS_RSA_EXPORT_WITH_RC4_128_SHA
    TLS_DHE_RSA_EXPORT_WITH_DES40_CBC_SHA
    TLS_DHE_DSS_EXPORT_WITH_DES40_CBC_SHA
    TLS_ECDHE_RSA_EXPORT_WITH_AES_256_CBC_SHA
    TLS_ECDH_RSA_EXPORT_WITH_AES_256_CBC_SHA
    TLS_DHE_RSA_EXPORT_WITH_AES_128_CBC_SHA
    TLS_DHE_RSA_WITH_AES_128_CBC_SHA
    TLS_RSA_EXPORT_WITH_RC4_128_SHA
    TLS_RSA_EXPORT_WITH_RC4_128_MD5
3.2. MESSAGE Transaction Over TLS

Once the TLS session is set up, the following MESSAGE request (as defined in [RFC3428] is sent from fluffy@example.com to kumiko@example.net. Note that the URI has a SIPS URL and that the VIA indicates that TLS was used. In order to format this document, the <allOneLine> convention from [RFC4475] is used to break long lines. The actual message does not contain the line breaks contained within those tags.
When a User Agent (UA) goes to send a message to example.com, the UA can see if it already has a TLS connection to example.com and if it does, it may send the message over this connection. A UA should have some scheme for reusing connections as opening a new TLS connection for every message results in awful performance. Implementers are encouraged to read [RFC5923] and [RFC3263].

The response is sent from example.net to example.com over the same TLS connection. It is shown below.

SIP/2.0 200 OK
Via: SIP/2.0/TLS 192.0.2.2:15001;branch=z9hG4bK-d8754z-c785a077a9a8451b-1---d8754z-;rport=50738
To: sips:kumiko@example.net:5061;tag=0d075510
From: sips:fluffy@example.com:15001;tag=1a93430b
Call-ID: OTZmMDE2OWNlYTVjNDkzYzBhMWRLMDU4NDEzMmU4ZTQ.
CSeq: 4308 MESSAGE
Content-Length: 0
4. Callflow with S/MIME-secured Message

4.1. MESSAGE Request with Signed Body

Below is an example of a signed message. The values on the Content-Type line (multipart/signed) and on the Content-Disposition line have been broken across lines to fit on the page, but they are not broken across lines in actual implementations.

MESSAGE sip:kumiko@example.net SIP/2.0
Via: SIP/2.0/TCP 192.0.2.2:15001;
    branch=z9hG4bK-d8754z-3a922b6dc0f0ff37-1---d8754z-;
    rport=50739
Max-Forwards: 70
To: <sip:kumiko@example.net>
From: <sip:fluffy@example.com>;tag=ef6bad5e
Call-ID: N2NiZjI0NjRjNDQ0MTY1NDRjNWNmMGU1MDRhYmI.
CSeq: 8473 MESSAGE
Accept: multipart/signed, text/plain, application/pkcs7-mime,
       application/sdp, multipart/alternative
Content-Type: multipart/signed;boundary=3b515e121b43a911;
    micalg=sha1;protocol="application/pkcs7-signature"
Content-Length: 774
--3b515e121b43a911
Content-Type: text/plain
Content-Transfer-Encoding: binary
Hello!
--3b515e121b43a911
Content-Type: application/pkcs7-signature;name=smime.p7s
Content-Disposition: attachment;handling=required;
    filename=smime.p7s
--3b515e121b43a911--
It is important to note that the signature ("BINARY BLOB 1") is computed over the MIME headers and body, but excludes the multipart boundary lines. The value on the Message-body line ends with CRLF. The CRLF is included in the boundary and is not part of the signature computation. To be clear, the signature is computed over data starting with the "C" in the "Content-Type" and ending with the "!
" in the "Hello!".

Content-Type: text/plain
Content-Transfer-Encoding: binary

Hello!

Following is the ASN.1 parsing of encrypted contents referred to above as "BINARY BLOB 1". Note that at address 30, the hash for the signature is specified as SHA-1. Also note that the sender’s certificate is not attached as it is optional in [RFC5652].

```
0  472: SEQUENCE {
  4  9:   OBJECT IDENTIFIER signedData (1 2 840 113549 1 7 2)
15  457:   [0] {
19  453:     SEQUENCE {
23   1:       INTEGER 1
26   11:       SET {
28   9:         SEQUENCE {
30   5:           OBJECT IDENTIFIER sha1 (1 3 14 3 2 26)
37   0:             NULL
 :         }
 :         }
39   11:       SEQUENCE {
41   9:         OBJECT IDENTIFIER data (1 2 840 113549 1 7 1)
 :         }
52  420:       SET {
56  416:         SEQUENCE {
60   1:           INTEGER 1
63  125:         SEQUENCE {
65  112:           SEQUENCE {
67   9:             SET {
71   3:               OBJECT IDENTIFIER countryName (2 5 4 6)
76   2:                 PrintableString 'US'
 :               }
 :             }
80   19:             SET {
82  17:               SEQUENCE {
84   3:                 OBJECT IDENTIFIER
 :                   stateOrProvinceName (2 5 4 8)
89   10:                 UTF8String 'California'
```

101 17: SET {
103 15:   SEQUENCE {
105  3:     OBJECT IDENTIFIER localityName (2 5 4 7)
110  8:     UTF8String 'San Jose'

120 14: SET {
122 12:   SEQUENCE {
124  3:     OBJECT IDENTIFIER
129  5:     UTF8String 'sipit'

136 41: SET {
138 39:   SEQUENCE {
140  3:     OBJECT IDENTIFIER
145 32:     UTF8String 'Sipit Test Certificate Authority'

179  9: INTEGER 00 96 A3 84 17 4E EF 8A 4D

190  9: SEQUENCE {
192  5:   OBJECT IDENTIFIER sha1 (1 3 14 3 2 26)
199  0:   NULL

201 13: SEQUENCE {
203  9:   OBJECT IDENTIFIER
214  0:   NULL

216 256: OCTET STRING
74 4D 21 39 D6 E2 E2 2C 30 5A AA BC 4E 60 8D 69
A7 E5 79 50 1A B1 7D 4A D3 C1 03 9F 19 7D A2 76
97 B3 CE 30 CD 62 4B 96 20 35 DB C1 64 D9 33 92
96 CD 28 03 98 6E 2C 0C F6 8D 93 40 F2 88 DA 29
AD 0B C2 0E F9 D3 6A 95 2C 79 6E C2 3D 62 E6 54
A9 1B AC 66 DB 16 B7 44 6C 03 1B 71 9C EE C9 EC
4D 93 B1 CF F5 17 79 C5 C8 BA 2F A7 6C 4B DC CF
62 A3 F3 1A 1B 24 E4 40 66 3C 4F 87 86 BF 09 6A
7A 43 60 2B FC D8 3D 2B 57 17 CB 81 03 2A 56 69
81 82 FA 7B 8E D2 3A 2F FA A3 C5 EA 8B E8 0C 36
1B BC DC FD 1B 8C 2E 0F 01 AF D9 E1 04 0E 4E 50
94 75 7C BD D9 0B DD AA FA 36 E3 EC E4 A5 35 46
SHA-1 parameters may be omitted entirely, instead of being set to NULL, as mentioned in [RFC3370]. The above dump of Blob 1 has SHA-1 parameters set to NULL. Below are the same contents signed with the same key, but omitting the NULL according to [RFC3370]. This is the preferred encoding. This is covered in greater detail in Section 5.

```
0  468: SEQUENCE {
  4  9:   OBJECT IDENTIFIER signedData (1 2 840 113549 1 7 2)
15  453:     [0] {
19  449:      SEQUENCE {
23   1:       INTEGER 1
26   9:       SET {
28   7:        SEQUENCE {
30   5:         OBJECT IDENTIFIER sha1 (1 3 14 3 2 26)
37   11:        SEQUENCE {
39   9:         OBJECT IDENTIFIER data (1 2 840 113549 1 7 1)
50  418:        SET {
54  414:         SEQUENCE {
58   1:           INTEGER 1
61  125:          SEQUENCE {
63  112:            SEQUENCE {
67   9:             SEQUENCE {
69   3:              OBJECT IDENTIFIER countryName (2 5 4 6)
74   2:              PrintableString 'US'
78  19:                SET {
80  17:                  SEQUENCE {
82   3:                   OBJECT IDENTIFIER
85  11:                    stateOrProvinceName (2 5 4 8)
87  10:                    UTF8String 'California'
```

99 17:
  SET {
101 15:
    SEQUENCE {
103 3:
      OBJECT IDENTIFIER localityName (2 5 4 7)
108 8:
        UTF8String 'San Jose'
    : }
  : }
118 14:
  SET {
120 12:
    SEQUENCE {
122 3:
      OBJECT IDENTIFIER

127 5:
        organizationName (2 5 4 10)
    : }
  : }
134 41:
  SET {
136 39:
    SEQUENCE {
138 3:
      OBJECT IDENTIFIER

143 32:
        organizationalUnitName (2 5 4 11)
    : }
  : }
177 9:
  INTEGER 00 96 A3 84 17 4E EF 8A 4D

188 7:
  SEQUENCE {
190 5:
    OBJECT IDENTIFIER sha1 (1 3 14 3 2 26)
  : }
197 13:
  SEQUENCE {
199 9:
    OBJECT IDENTIFIER

210 0:
    rsaEncryption (1 2 840 113549 1 1 1)
  : }
212 256:
  OCTET STRING

  74 4D 21 39 D6 E2 E2 2C 30 5A AA BC 4E 60 8D 69
  : A7 E5 79 50 1A B1 7D 4A D3 C1 03 9F 19 7D A2 76
  : 97 B3 CE 30 CD 62 4B 96 20 35 DB C1 64 D9 33 92
  : 96 CD 28 03 98 6E 2C 0C F6 BD 93 40 F2 88 DA 29
  : AD 0B C2 0E F9 D3 6A 95 2C 79 6E C2 3D 62 E6 54
  : A9 1B AC 66 DB 16 B7 44 6C 03 1B 71 9C EE C9 EC
  : 4D 93 B1 CF F5 17 79 C5 C8 BA 2F A7 6C 4B DC CF
  : 62 A3 F3 1A 1B 24 E4 40 66 3C 4F 87 86 BF 09 6A
  : 7A 43 60 2B FC D8 3D 2B 57 17 CB 81 03 2A 56 69
  : 81 82 FA 78 DE D2 3A 2F FA A3 C5 EA 8B E8 0C 36
  : 1B BC DC FD 1B 8C 2E 0F 01 AF D9 E1 04 0E 4E 50
  : 94 75 7C BD D9 0B DD AA FA 36 E3 EC E4 A5 35 46
  : BE A2 97 1D AD BA 44 54 3A ED 94 DA 76 4A 51 BA
  : A4 7D 7A 62 BF 2A 2F F2 5C 5A FE CA E6 B9 DC 5D
  : EA 26 F2 35 17 19 20 CE 97 96 4E 72 9C 72 FD 1F
4.2. MESSAGE Request with Encrypted Body

Below is an example of an encrypted text/plain message that says "Hello!". The binary encrypted contents have been replaced with the block "BINARY BLOB 2".

```plaintext
MESSAGE sip:kumiko@example.net SIP/2.0
Via: SIP/2.0/TCP 192.0.2.2:15001;
branch=z9hG4bK-d8754z-c276232b541dd527-1---d8754z--;
   rport=50741
Max-Forwards: 70
To: <sip:kumiko@example.net>
From: <sip:fluffy@example.com>;tag=7a2e3025
Call-ID: MDYyMDhhODA3NWE2ZjEyYzAwOTZlMjExNWI2ZWQwZGM.
CSeq: 3260 MESSAGE
Accept: multipart/signed, text/plain, application/pkcs7-mime,
   application/sdp, multipart/alternative
Content-Disposition: attachment;handling=required;
   filename=smime.p7m
Content-Transfer-Encoding: binary
Content-Type: application/pkcs7-mime;smime-type=enveloped-data;
   name=smime.p7m
Content-Length: 565

**************
* BINARY BLOB 2 *
**************

Following is the ASN.1 parsing of "BINARY BLOB 2". Note that at
address 454, the encryption is set to aes128-CBC.

0  561: SEQUENCE {

4    9:   OBJECT IDENTIFIER envelopedData (1 2 840 113549 1 7 3)
15   546:   [0] {
19   542:     SEQUENCE {
23    1:       INTEGER 0
26   409:       SET {
30   405:         SEQUENCE {
34    1:           INTEGER 0
37   125:           SEQUENCE {
39  112:             SEQUENCE {
41   11:               SET {
43    9:                 SEQUENCE {
45    3:                   OBJECT IDENTIFIER countryName (2 5 4 6)
48    2:                   PrintableString 'US'
51    1:                   };
53    8:                   };
57   17:                 SET {
59   15:                   OBJECT IDENTIFIER stateOrProvinceName (2 5 4 8)
62   10:                   UTF8String 'California'
65    1:                   };
68   17:                 SET {
70   15:                   OBJECT IDENTIFIER localityName (2 5 4 7)
73    8:                   UTF8String 'San Jose'
76    1:                   };
79   17:                 SET {
82   15:                   OBJECT IDENTIFIER organizationName (2 5 4 10)
85    5:                   UTF8String 'sipit'
88    1:                   };
91   17:                 SET {
94   15:                   OBJECT IDENTIFIER organizationalUnitName (2 5 4 11)
97   32:                   UTF8String 'Sipit Test Certificate Authority'
100    1:                   };
103   41:                 SET {
106   39:                   SEQUENCE {
109   3:                     OBJECT IDENTIFIER
112   3:                      .organizationName (2 5 4 10)
115   5:                       UTF8String 'sipit'
118    1:                       };
121   39:                     SEQUENCE {
124   3:                       OBJECT IDENTIFIER
127   3:                           .organizationUnitName (2 5 4 11)
130   32:                           UTF8String 'Sipit Test Certificate Authority'
133    1:                           };
136   39:                     SEQUENCE {
139   3:                       OBJECT IDENTIFIER
142   3:                           .organizationName (2 5 4 10)
145   5:                           UTF8String 'sipit'
148    1:                           };
151   39:                     SEQUENCE {
154   3:                       OBJECT IDENTIFIER
157   3:                           .organizationUnitName (2 5 4 11)
160   32:                           UTF8String 'Sipit Test Certificate Authority'
163    1:                           };
166    9:                         OBJECT IDENTIFIER
168    9:                             };
171    8:                             INTEGER 00 96 A3 84 17 4E EF 8A 4E
174    1:                             };
177   13:                           SEQUENCE {
180   9:                             OBJECT IDENTIFIER
183   9:                               };
186   8:                               INTEGER 00 96 A3 84 17 4E EF 8A 4E
4.3. MESSAGE Request with Encrypted and Signed Body

In the example below, some of the header values have been split across multiple lines. Where the lines have been broken, the <allOneLine> convention has been used. This was only done to make it fit in the RFC format. Specifically, the application/pkcs7-mime
Content-Type line is one line with no whitespace between the "mime;" and the "smime-type". The values are split across lines for formatting, but are not split in the real message. The binary encrypted content has been replaced with "BINARY BLOB 3", and the binary signed content has been replaced with "BINARY BLOB 4".
MESSAGE sip:kumiko@example.net SIP/2.0
<allOneLine>
Via: SIP/2.0/TCP 192.0.2.2:15001;
     branch=z9hG4bK-d8754z-97a26e59b7262b34-1---d8754z-;
     rport=50742
</allOneLine>
Max-Forwards: 70
To: <sip:kumiko@example.net>
From: <sip:fluffy@example.com>;tag=379f5b27
Call-ID: MjYwMzdjYTY3YWRkYzgzMjU0MGI4Mzc2Njk1YzJlNzE.
CSeq: 5449 MESSAGE
<allOneLine>
Accept: multipart/signed, text/plain, application/pkcs7-mime,
        application/sdp, multipart/alternative
</allOneLine>
<allOneLine>
Content-Type: multipart/signed;boundary=e8df6e1ce5d1e864;
              micalg=sha1;protocol="application/pkcs7-signature"
</allOneLine>
Content-Length: 1455
--e8df6e1ce5d1e864
<allOneLine>
Content-Type: application/pkcs7-mime;smime-type=enveloped-data;
              name=smime.p7m
</allOneLine>
<allOneLine>
Content-Disposition: attachment;handling=required;
                     filename=smime.p7
</allOneLine>
Content-Transfer-Encoding: binary

***************
* BINARY BLOB 3 *
***************
--e8df6e1ce5d1e864
Content-Type: application/pkcs7-signature;name=smime.p7s
<allOneLine>
Content-Disposition: attachment;handling=required;
                     filename=smime.p7s
</allOneLine>
Content-Transfer-Encoding: binary

***************
* BINARY BLOB 4 *
***************
--e8df6e1ce5d1e864--
Below is the ASN.1 parsing of "BINARY BLOB 3".

```
0 561: SEQUENCE {
  4   9:   OBJECT IDENTIFIER envelopedData (1 2 840 113549 1 7 3)
15 546:   [0] {
  23  1:     INTEGER 0
  26  409:       SET {
  30  405:         SEQUENCE {
  34  1:           INTEGER 0
  37  125:           SEQUENCE {
  39  112:             SEQUENCE {
  41   11:               SET {
  43    9:                 SEQUENCE {
  45    3:                   OBJECT IDENTIFIER countryName (2 5 4 6)
  50    2:                   PrintableString 'US'
  54   19:                 }
  56   17:               SEQUENCE {
  58    3:                 OBJECT IDENTIFIER stateOrProvinceName (2 5 4 8)
  63   10:                 UTF8String 'California'
  69   17:               SEQUENCE {
  75   17:                 SET {
  77   15:                 SEQUENCE {
  79    3:                   OBJECT IDENTIFIER localityName (2 5 4 7)
  84   8:                   UTF8String 'San Jose'
  89   14:                 }
  94   17:               }
  96   12:               SEQUENCE {
  98    3:                 OBJECT IDENTIFIER organizationName (2 5 4 10)
 103    5:                 UTF8String 'sipit'
 108    5:                 }
 113   41:               SET {
 117   39:                 SEQUENCE {
 121    3:                   OBJECT IDENTIFIER organizationalUnitName (2 5 4 11)
 126   32:                   UTF8String 'Sipit Test Certificate Authority'
 135    3:                   }
 138   17:               }
 142   9:               INTEGER 00 96 A3 84 17 4E EF 8A 4E

:  }
  166 9:  OBJECT IDENTIFIER
  :     rsaEncryption (1 2 840 113549 1 1 1)
  177 0:  NULL
  :
  179 256:  OCTET STRING
  :     49 11 0B 11 52 A9 9D E3 AA FB 86 CB EB 12 CC 8E
  :     96 9D 85 3E 80 D2 7C C4 9B B7 81 4B B5 FA 13 80
  :     6A 6A B2 34 72 D8 C0 82 60 DA B3 43 F8 51 8C 32
  :     8B DD D0 76 6D 9C 46 73 C1 44 A0 10 FF 16 A4 83
  :     74 85 21 74 7D E0 FD 42 C0 97 00 82 A2 80 81 22
  :     9C A2 82 0A 85 F0 6B EF 9A D7 6D 1D 24 2B A9 5E
  :     B3 9A A0 3E A7 D9 1D 1C D7 42 CB 6F A5 81 66 23
  :     28 00 7C 99 6A B6 03 3F 7E F6 48 EA 91 49 35 F1
  :     FD 40 54 5D AC F7 84 EA 3F 27 43 FD DE E2 10 DD
  :     63 C4 35 4A 13 63 0B 6D 0D 9A D5 AB 72 39 69 8C
  :     65 4C 44 C4 A3 31 60 79 B9 A8 A3 A1 03 FD 41 25
  :     12 E5 F3 F8 47 CE 8C 42 D9 26 77 A5 57 AF 1A 95
  :     BF 05 A5 E9 47 F2 D1 AE DC 13 7E 1B 83 5C 8C C4
  :     1F 31 BC 59 E6 FD 6E 9A B0 91 EC 71 A6 7F 28 3E
  :     23 1B 40 E2 C0 60 CF 5E 5B 86 08 62 B4 B7 DB
  :     00 DD AC 3A 39 27 E2 7C 96 AD 8A E9 C3 B8 06 5E
  :
  439 124:  SEQUENCE {
  452 29:  OBJECT IDENTIFIER data (1 2 840 113549 1 7 1)
  454 9:  SEQUENCE {
  465 16:  OCTET STRING
  :     88 9B 13 75 A7 66 14 C3 CF CD C6 FF D2 91 5D A0
  :
  483 80:  [0]
  :     80 0B A3 B7 57 89 B4 F4 70 AE 1D 14 A9 35 DD F9
  :     1D 66 29 46 52 40 13 E1 3B 4A 23 E5 EC AB F9 35
  :     A6 B6 A4 BE C0 02 31 06 19 C4 39 22 7D 10 4C 0D
  :     F4 96 04 78 11 85 4E 7E E3 C3 BC B2 DF 55 17 79
  :     5F F2 4E E5 25 42 37 45 39 5D F6 DA 57 9A 4E 0B
  :
  :

Below is the ASN.1 parsing of "BINARY BLOB 4".

0 472: SEQUENCE {

4   9:   OBJECT IDENTIFIER signedData (1 2 840 113549 1 7 2)  
15  457:   [0] {  
19  453:     SEQUENCE {  
23   1:       INTEGER 1  
26  11:       SET {  
28   9:         SEQUENCE {  
30   5:           OBJECT IDENTIFIER sha1 (1 3 14 3 2 26)  
37  0:             NULL  
39   11:         SEQUENCE {  
41   9:           OBJECT IDENTIFIER data (1 2 840 113549 1 7 1)  
44   1:             NULL  
52  420:         SET {  
56  416:           SEQUENCE {  
60   1:             INTEGER 1  
63  125:           SEQUENCE {  
65  112:             SEQUENCE {  
67   11:               SET {  
69    9:                 SEQUENCE {  
71    3:                   OBJECT IDENTIFIER countryName (2 5 4 6)  
76    2:                     PrintableString 'US'  
78    0:                 }  
80    19:               SET {  
82    17:                 SEQUENCE {  
84    3:                   OBJECT IDENTIFIER : stateOrProvinceName (2 5 4 8)  
89   10:                     UTF8String 'California'  
92   0:                 }  
98  101:               SET {  
101  17:                 SEQUENCE {  
103  15:                   OBJECT IDENTIFIER localityName (2 5 4 7)  
108   8:                     UTF8String 'San Jose'  
111   0:                 }  
117   0:               }  
123  120:             SET {  
125  12:             SEQUENCE {  
127  3:               OBJECT IDENTIFIER : organizationName (2 5 4 10)  
131  5:                 UTF8String 'sipit'  
134   0:               }  
138  136:             SET {  
139  41:             SEQUENCE {  
140  3:                OBJECT IDENTIFIER : organizationalUnitName (2 5 4 11)  
144  39:                UTF8String 'sipit engineers'  
148   0:             }  
154  140:           }  
157  125:         }  
159  112:       }  
162  11:     }  
164  9:   }  
166  3:   }  
168  2:   }  
171   0: }  
196  453: }  
200  9: {
204  0:  
UTF8String 'Sipit Test Certificate Authority'

SEQUENCE {
    OBJECT IDENTIFIER shal (1 3 14 3 2 26)
    NULL
}

SEQUENCE {
    OBJECT IDENTIFIER rsaEncryption (1 2 840 113549 1 1 1)
    NULL
}

OCTET STRING

6E 51 AC 24 2E BA 7C A1 EE B0 A8 55 BC D4 64 5D
: E5 29 09 5F B2 AF AA 6F 91 D2 97 79 32 5B AF CA
: FE A1 73 FC E5 57 4E C6 3B 67 35 AA E4 78 1E 59
: 93 EE 67 63 77 1E 7A 82 BC 1E 26 0F 39 75 0C A6
: 26 92 01 6A B7 5D F0 C0 2C 51 46 FB A7 36 44 E3
: 64 C6 11 CB 0B 6B FD F3 6D 7C FD 3E AE 2E 91 BB
: 78 9E F4 1B A1 20 68 B9 DE D3 E3 0C FC F7 14 9A
: 2C 64 AB 27 52 BD 52 EC 27 B8 14 BD DB C3 54 C7
: EA 48 DB 07 E9 9B 2E C8 BE 62 A2 76 83 53 37 E8
: 02 4B D1 86 E9 DF 2E BD 93 39 EC 2F 01 53 A0 7F
: 1A B9 A6 31 FC E7 91 1C DB 22 4A 67 83 94 B2 4E
: 28 A9 CD DE 4A 04 6A E0 86 90 7B 58 5F DB 7A 96
: 96 A0 25 61 C2 58 A2 28 E5 B3 B2 F1 6D 51 06 9C
: 78 61 0D D8 3A A7 9F A3 B5 87 0B 80 11 C2 A9 1A
: E5 17 1C EB 82 55 AB CD 04 E7 D9 5B 11 E8 B7 47
: FE FD CC B7 DB 47 6F 77 85 9E 24 D8 11 E1 E4 7D

5. Observed Interoperability Issues

This section describes some common interoperability problems. These were observed by the authors at SIPit interoperability events. Implementers should be careful to verify that their systems do not introduce these common problems, and, when possible, make their clients forgiving in what they receive. Implementations should take extra care to produce reasonable error messages when interacting with software that has these problems.

Some SIP clients incorrectly only do SSLv3 and do not support TLS. See Section 26.2.1 of [RFC3261].

Many SIP clients were found to accept expired certificates with no warning or error. See Section 4.1.2.5 of [RFC5280].

When used with SIP, TLS and S/MIME provide the identity of the peer that a client is communicating with in the Subject Alternative Name in the certificate. The software checks that this name corresponds to the identity the server is trying to contact. Normative text describing path validation can be found in Section 7 of [RFC5922] and Section 6 of [RFC5280]. If a client is trying to set up a TLS connection to good.example.com and it gets a TLS connection set up with a server that presents a valid certificate but with the name evil.example.com, it will typically generate an error or warning of some type. Similarly with S/MIME, if a user is trying to communicate with sip:fluffy@example.com, one of the items in the Subject Alternate Name set in the certificate will need to match according to the certificate validation rules in Section 23 of [RFC3261] and Section 6 of [RFC5280].

Some implementations used binary MIME encodings while others used base64. It is advisable that implementations send only binary and are prepared to receive either. See Section 3.2 of [RFC5621].

In several places in this document, the messages contain the encoding for the SHA-1 digest algorithm identifier. The preferred form for encoding as set out in Section 2 of [RFC3370] is the form in which the optional AlgorithmIdentifier parameter field is omitted. However, [RFC3370] also says the recipients need to be able to receive the form in which the AlgorithmIdentifier parameter field is present and set to NULL. Examples of the form using NULL can be found in Section 4.2 of [RFC4134]. Receivers really do need to be able to receive the form that includes the NULL because the NULL form, while not preferred, is what was observed as being generated by most implementations. Implementers should also note that if the algorithm is MD5 instead of SHA-1, then the form that omits the AlgorithmIdentifier parameters field is not allowed and the sender
has to use the form where the NULL is included.

The preferred encryption algorithm for S/MIME in SIP is AES as defined in [RFC3853].

Observed S/MIME interoperability has been better when UAs did not attach the senders’ certificates. Attaching the certificates significantly increases the size of the messages, which should be considered when sending over UDP. Furthermore, the receiver cannot rely on the sender to always send the certificate, so it does not turn out to be useful in most situations.

Please note that the certificate path validation algorithm described in Section 6 of [RFC5280] is a complex algorithm for which all of the details matter. There are numerous ways in which failing to precisely implement the algorithm as specified in Section 6 of [RFC5280] can create a security flaw, a simple example of which is the failure to check the expiration date that is already mentioned above. It is important for developers to ensure that this validation is performed and that the results are verified by their applications or any libraries that they use.
6. Additional Test Scenarios

This section provides a non-exhaustive list of tests that implementations should perform while developing systems that use S/MIME and TLS for SIP.

Much of the required behavior for inspecting certificates when using S/MIME and TLS with SIP is currently underspecified. The non-normative recommendations in this document capture the current folklore around that required behavior, guided by both related normative works such as [RFC4474] (particularly, Section 13.4 Domain Names and Subordination) and informative works such as [RFC2818] Section 3.1. To summarize, test plans should:

- For S/MIME secured bodies, assure that the peer’s URI (address-of-record, as per [RFC3261] Section 23.3) appears in the subjectAltName of the peer’s certificate as a uniformResourceIdentifier field.

- For TLS, assure that the peer’s hostname appears as described in [RFC5922]. Also:
  * assure an exact match in a dNSName entry in the subjectAltName if there are any dNSNames in the subjectAltName. Wildcard matching is not allowed against these dNSName entries. See Section 7.1 of [RFC5922].
  * assure that the most specific CommonName in the Subject field matches if there are no dNSName entries in the subjectAltName at all (which is not the same as there being no matching dNSName entries). This match can be either exact, or against an entry that uses the wildcard matching character ‘*’.

The peer’s hostname is discovered from the initial DNS query in the server location process [RFC3263].

- IP addresses can appear in subjectAltName ([RFC5280]) of the peer’s certificate, e.g. "IP:192.168.0.1". Note that if IP addresses are used in subjectAltName, there are important ramifications regarding the use of Record-Route headers that also need to be considered. See Section 7.5 of [RFC5922]. Use of IP addresses instead of domain names is inadvisable.

For each of these tests, an implementation will proceed past the verification point only if the certificate is "good". S/MIME protected requests presenting bad certificate data will be rejected. S/MIME protected responses presenting bad certificate information will be ignored. TLS connections involving bad certificate data will
not be completed.

1. S/MIME : Good peer certificate
2. S/MIME : Bad peer certificate (peer URI does not appear in subjectAltName)
3. S/MIME : Bad peer certificate (valid authority chain does not end at a trusted CA)
4. S/MIME : Bad peer certificate (incomplete authority chain)
5. S/MIME : Bad peer certificate (the current time does not fall within the period of validity)
6. S/MIME : Bad peer certificate (certificate, or certificate in authority chain, has been revoked)
7. S/MIME : Bad peer certificate ("Digital Signature" is not specified as an X509v3 Key Usage)
8. TLS : Good peer certificate (hostname appears in dNSName in subjectAltName)
9. TLS : Good peer certificate (no dNSNames in subjectAltName, hostname appears in CN of Subject)
10. TLS : Good peer certificate (CN of Subject empty, and subjectAltName extension contains an iPAddress stored in the octet string in network byte order form as specified in RFC 791 [RFC0791])
11. TLS : Bad peer certificate (no match in dNSNames or in the Subject CN)
12. TLS : Bad peer certificate (valid authority chain does not end at a trusted CA)
13. TLS : Bad peer certificate (incomplete authority chain)
14. TLS : Bad peer certificate (the current time does not fall within the period of validity)
15. TLS : Bad peer certificate (certificate, or certificate in authority chain, has been revoked)
16. TLS : Bad peer certificate ("TLS Web Server Authentication" is not specified as an X509v3 Key Usage)
17. TLS : Bad peer certificate (Neither "SIP Domain" nor "Any Extended Key Usage" specified as an X509v3 Extended Key Usage, and X509v3 Extended Key Usage is present)
7. IANA Considerations

No IANA actions are required.
8. Acknowledgments

Many thanks to the developers of all the open source software used to create these call flows. This includes the underlying crypto and TLS software used from openssl.org, the SIP stack from www.resiprocate.org, and the SIMPLE IMPP agent from www.sipimp.org. The TLS flow dumps were done with SSLDump from http://www.rtfm.com/ssldump. The book "SSL and TLS" [EKR-TLS] was a huge help in developing the code for these flows. It’s sad there is no second edition.

Thanks to Jim Schaad, Russ Housley, Eric Rescorla, Dan Wing, Tat Chan, and Lyndsay Campbell who all helped find and correct mistakes in this document.

Vijay Gurbani and Alan Jeffrey contributed much of the additional test scenario content.
9. Security Considerations

Implementers must never use any of the certificates provided in this document in anything but a test environment. Installing the CA root certificates used in this document as a trusted root in operational software would completely destroy the security of the system while giving the user the impression that the system was operating securely.

This document recommends some things that implementers might test or verify to improve the security of their implementations. It is impossible to make a comprehensive list of these, and this document only suggests some of the most common mistakes that have been seen at the SIPit interoperability events. Just because an implementation does everything this document recommends does not make it secure.

This document does not show any messages to check certificate revocation status (see Sections 3.3 and 6.3 of [RFC5280]) as that is not part of the SIP call flow. The expectation is that revocation status is checked regularly to protect against the possibility of certificate compromise or repudiation. For more information on how certificate revocation status can be checked, see [RFC2560] (Online Certificate Status Protocol) and [RFC5055] (Server-Based Certificate Validation Protocol).
10. Changelog

(RFC Editor: remove this section)

-02 to -03

* Re-worded "should" and "must" so that the document doesn’t sound like it is making normative statements. Actual normative behavior is referred to in the respective RFCs.

* Section 5: re-worded paragraphs 4 and 5 regarding subjectAltName, and added references.

* Section 6: added references, clarified use of IP addresses, and clarified which From/To URI is used for comparison (from section 23.2). Added an EKU test case.

* Section 9: added text about certificate revocation checking.

* Appendix B.3: new section to present certificate chains longer than 2 (non-root CA).

* Made examples consistently use <allOneLine> convention.

* CSeq looks more random.

* Serial numbers in certificates are non-zero.

* All flows re-generated using new certificates. IP addresses conform to RFC 5737.

* Updated references.

-01 to -02

* Draft is now informational, not standards track. Normative-sounding language and references to RFC 2119 removed.

* Add TODO: change "hello" to "Hello!" in example flows for consistency.

* Add TODO: Fix subjectAltName DNS:com to DNS:example.com and DNS:net to DNS:example.net.

* Add TODO: use allOneLine convention from RFC4475.

* Section 3: updated open issue regarding contact headers in MESSAGE.
* Section 3.2: added some text about RFC 3263 and connection reuse and closed open issue.

* Section 5: clarified text about sender attaching certs, closed issue.

* Section 5: clarified text about observed problems, closed issue.

* Section 5: closed issue about clients vs. servers vs. proxies.

* Section 6: updated section text and open issue where IP address is in subjectAltName.

* Section 6: added normative references and closed "folklore" issue.

* Section 6: added cases about cert usage and broken chains, updated OPEN ISSUE: we need a SIP EKU example.

* References: updated references to drafts and re-categorized informative vs. normative.

* Section 9: added some text about revocation status and closed issue.

* Appendix B: open issue: do we need non-root-CA certs and host certs signed by them for help in testing cases in Section 6?

* Miscellaneous minor editorial changes.

-00 to -01

* Addition of OPEN ISSUES.

* Numerous minor edits from mailing list feedback.

-00

* Changed RFC 3369 references to RFC 3852.

* Changed draft-ietf-sip-identity references to RFC 4474.

* Added an ASN.1 dump of CMS signed content where SHA-1 parameters are omitted instead of being set to ASN.1 NULL.

* Accept headers added to messages.
* User and domain certificates are generated with EKU as specified in Draft SIP EKU.

* Message content that is shown is computed using certificates generated with EKU.

* Message dump archive returned.

* Message archive contains messages formed with and without EKU certificates.

prior to -00

* Incorporated the Test cases from Vijay Gurbani’s and Alan Jeffrey’s Use of TLS in SIP draft

* Began to capture the folklore around where identities are carried in certificates for use with SIP

* Removed the message dump archive pending verification (will return in -02)
11. References

11.1. Normative References


[ RFC 5621 ] Camarillo, G., "Message Body Handling in the Session

11.2. Informative References


[ssldump-manpage]
Rescorla, E., "SSLDump manpage".
Appendix A. Making Test Certificates

These scripts allow you to make certificates for test purposes. The certificates will all share a common CA root so that everyone running these scripts can have interoperable certificates. WARNING - these certificates are totally insecure and are for test purposes only. All the CA created by this script share the same private key to facilitate interoperability testing, but this totally breaks the security since the private key of the CA is well known.

The instructions assume a Unix-like environment with openssl installed, but openssl does work in Windows too. OpenSSL version 0.9.8j was used to generate the certificates used in this document. Make sure you have openssl installed by trying to run "openssl". Run the makeCA script found in Appendix A.1; this creates a subdirectory called demoCA. If the makeCA script cannot find where your openssl is installed you will have to set an environment variable called OPENSSLDIR to whatever directory contains the file openssl.cnf. You can find this with a "locate openssl.cnf". You are now ready to make certificates.

To create certificates for use with TLS, run the makeCert script found in Appendix A.2 with the fully qualified domain name of the proxy you are making the certificate for. For example, "makeCert host.example.net domain eku". This will generate a private key and a certificate. The private key will be left in a file named domain_key_example.net.pem in Privacy Enhanced Mail (PEM) format. The certificate will be in domain_cert_example.net.pem. Some programs expect both the certificate and private key combined together in a Public-key Cryptography Standards (PKCS) #12 format file. This is created by the script and left in a file named example.net.p12. Some programs expect this file to have a .pfx extension instead of .p12 - just rename the file if needed. A file with a certificate signing request, called example.net.csr, is also created and can be used to get the certificate signed by another CA.

A second argument indicating the number of days for which the certificate should be valid can be passed to the makeCert script. It is possible to make an expired certificate using the command "makeCert host.example.net 0".

Anywhere that a password is used to protect a certificate, the password is set to the string "password".

The root certificate for the CA is in the file root_cert_fluffyCA.pem.

For things that need DER format certificates, a certificate can be
converted from PEM to DER with "openssl x509 -in cert.pem -inform PEM -out cert.der -outform DER".

Some programs expect certificates in PKCS #7 format (with a file extension of .p7c). You can convert these from PEM format to PKCS #7 with "openssl crl2pkcs7 -nocrl -certfile cert.pem -certfile demoCA/cacert.pem -outform DER -out cert.p7c"

IE (version 8), Outlook Express (version 6), and Firefox (version 3.5) can import and export .p12 files and .p7c files. You can convert a PKCS #7 certificate to PEM format with "openssl pkcs7 -in cert.p7c -inform DER -outform PEM -out cert.pem".

The private key can be converted to PKCS #8 format with "openssl pkcs8 -in a_key.pem -topk8 -outform DER -out a_key.p8c"

In general, a TLS client will just need the root certificate of the CA. A TLS server will need its private key and its certificate. These could be in two PEM files, a single file with both certificate and private key PEM sections, or a single .p12 file. An S/MIME program will need its private key and certificate, the root certificate of the CA, and the certificate for every other user it communicates with.

A.1. makeCA script

```
#!/bin/sh
set -x

rm -rf demoCA

mkdir demoCA
mkdir demoCA/certs
mkdir demoCA/crl
mkdir demoCA/newcerts
mkdir demoCA/private

# This is done to generate the exact serial number used for the RFC
echo "4902110184015C" > demoCA/serial
touch demoCA/index.txt

# You may need to modify this for where your default file is
# you can find where yours in by typing "openssl ca"
for D in /etc/ssl /usr/local/ssl /sw/etc/ssl /sw/share/ssl; do
    CONF=${OPENSSLDIR:=$D}/openssl.cnf
    [ -f $CONF ] && break
done

CONF=${OPENSSLDIR}/openssl.cnf
```
if [ ! -f $CONF ]; then
    echo "Can not find file $CONF - set your OPENSSLDIR variable"
    exit
fi

cp $CONF openssl.cnf

cat >> openssl.cnf

[ sipdomain_cert ]
subjectAltName=${ENV::ALTNAME}
basicConstraints=CA:FALSE
subjectKeyIdentifier=hash
authorityKeyIdentifier=keyid,issuer
keyUsage = nonRepudiation,digitalSignature,keyEncipherment
extendedKeyUsage=serverAuth,1.3.6.1.5.5.7.3.20

[ sipdomain_req ]
basicConstraints = CA:FALSE
subjectAltName=${ENV::ALTNAME}
subjectKeyIdentifier=hash

[ sipuser_cert ]
subjectAltName=${ENV::ALTNAME}
basicConstraints=CA:FALSE
subjectKeyIdentifier=hash
authorityKeyIdentifier=keyid,issuer
keyUsage = nonRepudiation,digitalSignature,keyEncipherment
extendedKeyUsage=emailProtection,1.3.6.1.5.5.7.3.20

[ sipuser_req ]
basicConstraints = CA:FALSE
subjectAltName=${ENV::ALTNAME}
subjectKeyIdentifier=hash

[ sipdomain_noeku_cert ]
subjectAltName=${ENV::ALTNAME}
basicConstraints=CA:FALSE
subjectKeyIdentifier=hash
authorityKeyIdentifier=keyid,issuer
keyUsage = nonRepudiation,digitalSignature,keyEncipherment

[ sipdomain_noeku_req ]
basicConstraints = CA:FALSE
subjectAltName=${ENV::ALTNAME}
subjectKeyIdentifier=hash

[ sipuser_noeku_cert ]
subjectAltName=${ENV::ALTNAME}

[ sipuser_noeku_req ]

basicConstraints=CA:FALSE
subjectKeyIdentifier=hash
authorityKeyIdentifier=keyid,issuer
keyUsage = nonRepudiation,digitalSignature,keyEncipherment

[ sipuser_noeku_req ]
basicConstraints = CA:FALSE
subjectAltName=$\{ENV::ALTNAME\}$
subjectKeyIdentifier=hash
EOF

cat > demoCA/private/cakey.pem <<EOF
-----BEGIN ENCRYPTED PRIVATE KEY-----
MIIFDjABGkqhkiG9w0BAQsAwIBAgIzAQG64yvKOEoqBvEfGaaZ9mlcni1d7tSBIR
3ucN2zGmdEUVuZvSAC29PuI07WBG565enDRszqj4WfqHxdA4gkJKgPEJebNlV
jxCMlJ9ykM2k5HDM8ILWuOccO/Rk+7999h13Qv1Ym7yCincorgd1TrAdmq127Y
QPGxio1lx6++6yixIDCv7Mwkydo31k9y/Tf2cz//dwuf/1Maag8HfsSN14Jqks
uAl41K5sCzPR1uUGdouOQSPG0CjgcskPcii6zvrI62KLPFvrgw5H9T9devQvC6O
VgllxbEG7I4vlzMy62/0LTQKL66bh8pszvvmmHjGo9s+f+pKJYygEHNEmRTm+
8M20xk6703h3sV6IC1DOAdR8sli1ThclmmM+i1rX9VIPpsdrzqJqVGYBBjJ7W8eQp
Ta411148bcB1tDDmss8ED4stPv/650A1jzAhBbpjP0W5T6f1cY+]i0q9vcunohcj+i
KVXsM34w0s6pMBjFQ+Awsw5bsIkEV11i0Yav1F7/BvF2sog3puM5W5y1chKLU2
ThjvMu0777Q0yQyJba0UK90zBVEvPNahrDINucbEkFrhN2psno7s4Uvrjik
uknKrm3gocDostyMXZ81Beyj06NhpcJH+boSvRok/d68aAsapy6qSG9hiJnNBd
itQ/f0+1o9MDujT/huj7ZFqdzNM3KAVsx0kmVM+GJBYke+cjxK6W801F91yCb
0pwpd+fgwFL2525FUfocjyvUfWkkxrvB+IMkv6sNdKcXHHzaxE6E6y169PvCaS1I
WNqEFhntblNzbeW4s3h8Ov12XVGCemaHkajSAhJKXCGpsXax2FntzpVFpRnpw
Yd9eml9xwge31aavRv6p6id051Zc7KjXv01CemPUT6RYbAMFNC0b7cqihe
kcqMkQoKwkEMEPyP2zdMhGzsSPKk10nn153grpy/8FT5dpuq7353yjKxKbAvl+kV
lrfz6B7/pURwLBC3tLB6G6Uu/Ese4ewUW+2UXJIAFopCcvWrUqk4j368e228dd
A27X4Z2+j+ysFNP10A7sH13A3m9zBqH3EEGsrDzkgk8qkmtl1hGq/8232EXueuf
81S00j0U08HAgkTtFwv5/0C0q6W9W9Smwzw7msZHoKMTQ8kz2LXXU wb6DBw5c6e/
UTUgVXgqhA8hmjsnVe9ftJDL66v9zlp4RvDrDzm4ITYybYh5uiqFbJDFLJnHnjho
TcmusH080Cxs64hlLRz2M460i+JSEpV7O7cHcWFN0VtWN908EK5CubnTnQn9VC
0Sky9R/ZwunA1LG3Z3B0uhWPyvvdNxlnQ3ie4t6cm1X1Eel4U2N0sPCOFF//NEn
BEc=
-----END ENCRYPTED PRIVATE KEY-----
EOF

cat > demoCA/cacert.pem <<EOF
-----BEGIN CERTIFICATE-----
# uncomment the following lines to generate your own key pair

```bash
# openssl req -newkey rsa:2048 -passin pass:password 
# -passout pass:password -set_serial 0x96a384174eef8a4c 
# -shal -x509 -keyout demoCA/private/cakey.pem 
# -out demoCA/cacert.pem -days 36500 -config $(CONF) <<EOF

# US
# California
# San Jose
# sipit
# Sipit Test Certificate Authority
#        
# EOF
```

 either randomly generate a serial number, or set it manually  
 hexdump -n 4 -e '4/1 "%04u"' /dev/random > demoCA/serial  
 echo 96a384174eef8a4d > demoCA/serial  

openssl crl2pkcs7 -nocrl -certfile demoCA/cacert.pem  

```
EOF
```
A.2. makeCert script

#!/bin/sh
set -x

# Make a symbolic link to this file called "makeUserCert"
# if you wish to use it to make certs for users.

# ExecName=$(basename $0)
#
# if [ $(ExecName) == "makeUserCert" ]; then
#   ExtPrefix="sipuser"
# elif [ $(ExecName) == "makeEkuUserCert" ]; then
#   ExtPrefix="sipuser_eku"
# elif [ $(ExecName) == "makeEkuCert" ]; then
#   ExtPrefix="sipdomain_eku"
# else
#   ExtPrefix="sipdomain"
# fi

if [ $# == 3 ]; then
  DAYS=36500
elif [ $# == 4 ]; then
  DAYS=$4
else
  echo "Usage: makeCert test.example.org user|domain eku|noeku [days]"
  echo "  makeCert alice@example.org [days]"
  echo "days is how long the certificate is valid"
  echo "days set to 0 generates an invalid certificate"
  exit 0
fi

ExtPrefix="sip"$2

if [ $3 == "noeku" ]; then
  ExtPrefix=${ExtPrefix}"_noeku"
fi

DOMAIN='echo $1 | perl -ne '(print "$1\n" if (/\w+/..*)$))/’

USER='echo $1 | perl -ne '{print "$1\n" if (/(^\w+)\0(\w+\..*)$/)}' ' 
ADDR=$1 

if [ $2 == "user" ]; then 
   CNVALUE=$USER 
else 
   CNVALUE=$DOMAIN 
fi 
rm -f ${ADDR}_*.pem 
rm -f ${ADDR}.p12 

case ${ADDR} in 
   *:*:) ALTNAME="URI:${ADDR}" ;; 
   *) ALTNAME="DNS:${DOMAIN},URI:sip:${ADDR}" ;; 
esac 
rm -f demoCA/index.txt 
touch demoCA/index.txt 
rm -f demoCA/newcerts/* 

export ALTNAME 

openssl genrsa -out ${ADDR}_key.pem 2048 
openssl req -new -config openssl.cnf -reqexts $(ExtPrefix)_req 
   -sha1 -key ${ADDR}_key.pem 
   -out ${ADDR}.csr -days ${DAYS} <<EOF 
US 
California 
San Jose 
sipit 
$(CNVALUE) 
EOF 

if [ $DAYS == 0 ]; then 
   openssl ca -extensions $(ExtPrefix)_cert -config openssl.cnf 
   -passin pass:password -policy policy Anything 
   -md sha1 -batch -notext -out ${ADDR}_cert.pem 
   -startdate 990101000000Z 
   -enddate 000101000000Z 
   -infiles ${ADDR}.csr 
else 

openssl ca -extensions ${ExtPrefix}_cert -config openssl.cnf \  
   -passin pass:password -policy policy_anything \  
   -md sha1 -days ${DAYS} -batch -notext -out ${ADDR}_cert.pem \  
   -infiles ${ADDR}.csr
fi

openssl pkcs12 -passin pass:password \  
   -passout pass:password -export \  
   -out ${ADDR}.p12 -in ${ADDR}_cert.pem \  
   -inkey ${ADDR}_key.pem -name ${ADDR} -certfile demoCA/cacert.pem

openssl x509 -in ${ADDR}_cert.pem -noout -text

case ${ADDR} in
   *@) mv ${ADDR}_key.pem user_key_${ADDR}.pem; \  
      mv ${ADDR}_cert.pem user_cert_${ADDR}.pem ;;
   *) mv ${ADDR}_key.pem domain_key_${ADDR}.pem; \  
      mv ${ADDR}_cert.pem domain_cert_${ADDR}.pem ;;
   esac
Appendix B. Certificates for Testing

This section contains various certificates used for testing in PEM format.

B.1. Certificates Using EKU

These certificates make use of the EKU specification described in [RFC5924].

Fluffy’s user certificate for example.com:

-----BEGIN CERTIFICATE-----
MIIEGTCCAwGgAwIBAgIJAJajhBdO74pNMA0GCSqGSIb3DQEBBQUAMHAxCzAAMBgN
BAYTA1VTMRmweQYDVQQIDApDYWxpZm9ybiwibhMrEwDwYDVQQHDhTYW4gSm9zZTEO
MAwGh1UECgwPc2lwaXQxKTAnBgNVBAsMiWNFcGl0IFRlc3QgQ2Vydm1maWNhdGUQ
QXV0aG9yaXR5MCAXDTExMDIwNzE5MzIxN1oYDzIxMTEwMTE0MTkzMjE3WjBWMQsw
CQYDVQQEwZmbwEwggEiMA0GCSqGSIb3DQEBAQUAA4IB道AwggEKAoIBAQCjLFkM6b
z7NOe+5kC7LE2orfTHU3D0araUL1fVQh3JH6k6BoMSi1ZJGWGcMi16td/aciiKg1r
2G9X37BFONYKBq0tjiKJu4N2tsnuJx3rkwEeDkYwnXnrctssz765e174TZQLAXe/6n
9a83pfjh7CN0HvBrcXgFo4Y/9V9k19xbcqcghVcwrKyuxR7FNuPSaAgP41gwYKYOIC0
TzzPoDrSl6C5XhBQu7ivjL8EanoaeeGiQTFeT5wEm01YNvbAbV+NnHPAHcyy0xjGzG
LRd6LkiQBG/rfthEgGXHUF4ahWL+yeWc0GRNCnckHM0JvdS+1prB8KZn493PtPLA
AgMBAAGjgc0wgcwUQYDVQR0BEBowSiYWC21lOmZsdWZmeUBLG6=FtxGxlLmNvbYY
VaW06zmlx1zm5Z0GV4YW1wbGUuY29thdwcvmZomZsdWzmeUBLG6=FtxGxlLmNvbTA
JBAgNVHkEAAjAMBGU1dDgQWB3FLwm0410U3J1rc3uORcuiQiz5iHUjAfBgNVHSME
DAWgBSVRX5fK+plmBR5Phjx21awBZ3JzALBgdNVHQ8EBAMCBeAwHQYDVVR0LBByWfAY
IKwYBBQUHAgQCCsGAQUBFwWUMA0GCSqGSIb3DQEBAQQUA41IAAcqY/YgU7I7f9Pv+
XN5j7uOByLKrji1uacV79IKpd2dPB8ujwvmafB8yke0+RK43w9xTDjeB9zRQvL5Is3
1dv0mmiuiUBL0bTC81wyDoENxwOhrRF9Xxl1RnuVvETzy/8i4P8OCgb1mDzLGN8Mfa
TrtCzFTPbDthRmH2br6/hEhMjHgbrb9x/XasVDuM1kQAOkNVYBQgEQ6SkQTnrBi0z
wbwLDlcvxpeSuJyPFaarWZyZnc3RuqjuRzgeyG4GgYUclVc2Bh1sONuBnLgH4w+9S
J8jMEOA4cONrhmo/vIMAgyY4lqtqwr4RLB4GRO4fvpqoldRLS3V1v28J
-----END CERTIFICATE-----

Fluffy’s private key for user certificate for example.com:

-----BEGIN PRIVATE KEY-----
MIIEvQIBADANBgkqhkiG9w0BAQEFAASCAIAA8Hqg9H5zS/QiU5GwKuL7p+6e67
fMv4HvSocaqFi+Z1oP5k5IyJ+2gkJ3KfSi1mEjDhYuOjovP07Yz6o7QZCJv4u
Tkm7jBzQa3nY6Uy77v5QbW69Q5RlDF8v7Rm3ld2nOs84DDXtuPfI1C6/wza7
-----END PRIVATE KEY-----

-----BEGIN RSA PRIVATE KEY-----
MIIEpAIBAAKCAQEy+aHVXQN4BX/zLhiKmzGCPce405FH3sVpKV06bcZx9g/Ic
1F/Pz1J9113E2z9NfQ6gPjJD1+DTRq2cncqCK3Krl1Fyu1EJUaCsk0pWtdQ4ox0
6n1o1hSK0Y0WVIT449FrtxQBT2EVRVgyGX0UHB/AMncPF0qeteJtZmP8V+CS7Hz
YpgI1WvAxzkOpaG5h8INLYVPfZl1dVdVtUgNZNQ9iWV/NW61g4N06T-R667ClNxK
AWRudO5YRvW3FdcgQR5WhaRXg0w2BNT9fe1El4akbSKqg8a+ueHaPQ5z9JmB+nT5
Ubb3gGYXSIqGwh4BmiqiKt/kmQ5y3Qsg6xUjvw5W5IDAQABoIAHAXMrGgRS0xWL
BPLbMk+ILSeRiR4+bywbg663SHTAB1Yzzv+w2Be2oMnv MJrEe0o4071266bJoZ
KrfCkmKqrYJKajXkrBW/jtZ6xCGWPCNA1lpx1IpIG5tD1gj8SAL0047hy0r3Ar
W0vuVQSYFX4BhdvZeszyRwCqnp3x0PSff95Ad+vujd5CYuFZCuYGkszQ3fi+Nia
Gqs01EuyolEv72rswe25+wtx3xqXBBZ4HXR+yQ9NbeE8ip2CWD1Uh1q1H18kw
V3oLkiIoV7V+m6Zx/uzaAMF0RDn5KET+b5DO1IkUAaa8LZsf59rvOKLqw7aAaj5
esXhAdGECgYE8A8930YqU2+AcEkjC5hygw1M/X5k/Icvz0pa0/in2hW71zGzg0A
FJjxuoIVxbsF9c7z/B6675Vw94emovLArqWhFaLfb2CsrleEHqQtGCuL30w6p
X0EebBF5Fm0K7qQ1Qgqj0NF5zP2JrsxGNrmFYFV9dcp/3Jp/I1ZEsCgYEALuqI
/718h9og1dmTPzmpvpsnA0RF/1u4X9AE41NRp09HjxOB7Vuat1ABtx09/ZNHn
LHZBTZ5R2R2RjbsXHZ3DfoSx91Q3r1x+TuEl4RcppHJkDkJvPDPnLoU0QbFG6
kyUtnEF+6V1IZvVnsMgB6wpHUL1cjdAZUl7ip6W4TECGyAMHpa7pExUD076rH9
tpeCsume541lsHtX0WboA0pVczuqesRdKmBW1jF7YoUs3yqH182JoPM8lamgf
9QJMzs9Yh
/5Y1aIuUJk+wQ9VnZJjmNM6OHdTVFQmE9VCEHL5/Mmxo6fiWZ8EjLJ7S7HVzz
yDqhtbh6WF5WYM15ZdX3xwKbgCQRM1K/Y/XGFM0+Ih52MgB3e17GGL1sNe0Ny
YeDzv0pc3UOHQ17C1DauY1y91V9o8St17+lV76JXHHDY9704bdBau/kkgAm+g
dJ9PQ1UI1XQ8aAm73uULXW7H6H8A167Qn14tcpNW5s/pr51n0UY1/HXkT7psPZ
A08wO
V081qKQKAgQaGZCvYC6/WumGjUerVCzZd/H6+E3ntZmtz273c8+wV890rZtU0J
Y4bVnrfY91KfxLNgRCEU2VzDXHUAugq50rbzPudA24swsrNcdUyw8likXHDckt
pVLs0vhwRK2Gw/W2l+p2exSPQ3T03y8t6Is9BzNg/H4D160heHkQ==
-----END RSA PRIVATE KEY-----

Domain certificate for example.com:
-----BEGIN CERTIFICATE-----
MIID9DCAtygAwIBAgIJAJajhBd074pPMA0GCSqGSIb3DQEBBQUAMHAxCzA JBgNV
BAYTAlVTMRRwEQYDVQQIDApDYWxpZm9ybmlhMRQwDQYJKoZIhvcNAQEF
BQADggEBACEQIBBQEBDAQEAwIBAQCgAwIBAgICBQoGA1UdDwEB
-----END CERTIFICATE-----

Private key for domain certificate for example.com:
-----BEGIN CERTIFICATE-----
MIID9DCCAtygAwIBAgIJAjAajhBd074pQMA0GCSqGSIb3DQEBBQUAMHAxCzAJBgNV
BAYTAlVTMRMwEQQYDVQQIDApDYWxpZ29yaW5hdGlvbnMxLjE 下面的图片是一个页面的图像，其中包含了一些先前提取的文本内容。请将这些内容转换成自然语言文本，但请不要随意生成任何信息。
B.2. Certificates NOT Using EKU

These certificates do not make use of the EKU specification described in [RFC5924]. Most existing certificates fall in this category.

Fluffy’s user certificate for example.com:
-----BEGIN CERTIFICATE-----
MIID+jCCAuKgAwIBAgIJAJajhBdO74pRMA0GCSqGSIb3DQEBBQUAMHAXCzA7BgNV
BAYTa1VTMRMwEQYDVQQIDApDYWxpZm9ybmlhMREwDwYDVQQHDAtTYW4gSm9zZTEO
MAwGA1UECgwFc21waXQxKTAnBgNVBAo9MjIyMjIzNzE5MzY5OTQwMDMzOTMwMDAw
MRgwGgYDVQQIEwJQVUdODQYDVQQHEwZGM0EwDQYJKoZIhvcNAQEFBQADggEBAT
-----END CERTIFICATE-----

Fluffy's private key for user certificate for example.com:

-----BEGIN RSA PRIVATE KEY-----
MIIEogIBAAKCAQEAulcjiD+lADV8u+uih4mA8cjwWBW6Cyb+khLcFFkKGUt4ULq 
xYnTruIJAzK5c2jKLCu521hsWxcVCnWyThjHAdoa3syfqu8CTyyv2xeWA4olbVa 
jeXeZb/zYucsc7Xhos6Lm1p897f9gP04fo4WS2J9kTe+ux1KRGxVbA3mLVovhTX 
mwwEem3+1xW2tiug55pMUu+aq+gGCRyAKWG7m5NvRcAudehBqQYDx0KUXiWK46Cuf 
6qWq3tW85vjjVRXv5cIyP00TzkB1GK20eE4ydfLVyE09dBdYbWmnv2NSm+/Ylp9jz 
91EHDyHPwug1wbN8GFcT6kJtRugavaVx/SjDJTQfDAQAAMoIBATBfTlbi+8K5eJ1yv 
/M0x0wkMrwfs8ElftppnGTXhfjJ31mBFIAS5hjD3GncDqamAL1kys6YEZ+mu/rnH 
wp2FXCO1FqSebd8TcM11bO27v0IXZUKTxR4aj41Y0YHrLq/g7yrfXJJer8WQ1KFMD 
FVKmLOWpk34+2jOOhqUDP3r3xhJClQ81Fc1hke2JoixNDoPfM3azTg8QPULQD2I 
mjww1IH1677G5c/6qMloOMOFeqv/3cUWIRmvFp4eyGHdNtuFXKFpB4DQMQMLTID8 
FoOHymHIOzSF+gygBFObo0YNgu2CqZrED9cfr0RortrbXf6mt+AklclfhhkFkaa 
JPSosbUCgYE4A4MaetKs7azhEYMc4TK0xhhV5H1i1jXr/6h+/yuF0O1OBjM9yu3 
5n6vlpyghNh2bKO80IWP0F4syyvKYR2elmUDraH29DKATRLEUK9U82RG4AmXmk 
G6ZsWoFx6Jf35OnAKVjt/7aN9jkcK1v6EFyQGYEXb4pi0fhhFbJBAec28CyEAI1dmx 
iKJ3d-jJW9ypNk51jYJ3r+a5qPNWnvjgGKQQe3Y6+r5lxmW0HzWxoCBOYRhwhHRA/ 
SxH9j38rECJnkxpeAoX2Gc0l5U6khHRw/rdf/3+SShQPrugAdIwNgkcu8BVPR 
8uP2CgJoDB15U2R97GVK98x8k2S6fDK32mQMCGbV/KH3R8VY7jiKcC21U0L 
J1E3/gb4S+wQ8YELth0FVPC0pDSluZld1tFwRofUraae01k/+HesSfiiJD1gNh6m 
cDPFO7/vh47zUWrZFCg95n4nk51hi1KNR1vn9e1JTudjLCSnpfC2jU3X1OBe+n 
k6zuFFoUuWFSc2iujibwKbgCt6RHejlkDe2FniX8r7D88y/W9wxVtDWqgiE4x 
XQ/0FPA81jBKTAq5gcsp2xbabDZPjc7Veta21A8FvQPVZCVsaA4413zVsO 
WYb1lT17ZXA2yu08wW/Gnds00u11TRGX6w+sAY0r11/M8k/tokn5HfEYsErbq 
Y/w3oaGAsj0C9fjy2baBH8sQaim/Rx3hFO4myOGWhxRmeeoO2Yd0M1dlr1z 
A/sQRvVoFHrwoayV2KakLprEGyxEqCdMms1h9xYAXcfw23RrFqC39DyB9RTRkwa 
ArJmcEdRES0sIYhXGf1QMgivj1UXMWeYcLtqQKW1LLDXYFQ=-----END RSA PRIVATE KEY-----

Kumiko’s user certificate for example.net:
-----BEGIN CERTIFICATE-----
MIID+jCCAwIBAgIJAJajhBdO74pSMA0GCSqGSIb3DQEBBQUAMHAAxChAjBqNV
BAYTAlVTMRmE4YDVQQIDApDQyExMSowMB4XDTExMDIwNzE5MzQzMjEwNTEyMjA4
MjA3NTYzMTc2NzExMjgwMjkwWjBMQswCQYDVQQGEwJVUzETMBEGA1UECBMKQ2Fz
QQDBCgUETMBAAMCAwEwDQYJKoZIhvcNAQEFBQADggEBAD9e571tfAgJnKhB6
-----END CERTIFICATE-----

Kumiko’s private key for user certificate for example.net:

-----BEGIN CERTIFICATE-----
MIID+jCCAwIBAgIJAJajhBdO74pSMA0GCSqGSIb3DQEBBQUAMHAAxChAjBqNV
BAYTAlVTMRmE4YDVQQIDApDQyExMSowMB4XDTExMDIwNzE5MzQzMjEwNTEyMjA4
MjA3NTYzMTc2NzExMjgwMjkwWjBMQswCQYDVQQGEwJVUzETMBEGA1UECBMKQ2Fz
QQDBCgUETMBAAMCAwEwDQYJKoZIhvcNAQEFBQADggEBAD9e571tfAgJnKhB6
-----END CERTIFICATE-----

-----BEGIN CERTIFICATE-----
MIID1TCCAr2gAwIBAgIJAJajhBdO74pTMA0GCSqGS1b3DQEBBQUAMHAxCzAJBgNV
BAYTA1VTMRMwEQYDVQQIDA0wYXwpZm9ybmlhMQswCQYDVQQGEwJVUzETMBEGA1UE
BxMEc3RvY3Jvc2xvcmQwHhcNMTgwMDExMDI0NTk0WhcNDgwMDExMDI0NTk0Whc
DAYDVQQIEwNjb3JuaWduIFRlc3QgU2VjdXJlIGV4cGFuY2V1cyBQb3JjdW50aW
QXV0aG9yaXR5MA0GCSqGSIb3DQEBCwYDAYDVQQJEwNsb3JpdGlvbiBBdXRob3
RoZIk1hvcgNAQEBQgA8EIoDCCAsQDQYBUBDaHlVw9CIgJzA6B0h8DQYJKoZIh
vcNAQEBBiMAKwggCSAgEIkDAABoAwIBAQBwDAgAAN5nAEMz/B9J/6p49t9fXx2
4mWVyc+d8ygGcFm8+Q7XH47o5nM+hX38iC4z18zP7nC+a7j4u7+x8n+0sWgj7k
-----END CERTIFICATE-----

Private key for domain certificate for example.com:
-----BEGIN CERTIFICATE-----
MIID1TCCAr2gAwIBAgIJAjAajhBdO74pUMA0GCSqGS1b3DQEBBQUAMHAxCzAJBgNV
BAYTAlVTMRwEQYDVQQIDApDYWxpZm9ybmlhMRERdWYDVQQHDAYW4gszZTEOMAwGA1UE
CgwFc2liwaXQxKTA0NCAgNVBAUwT3ExMTAeFw0xNzEwMTgwNzU5MDAwMFwXDTExMDIw
NzE5MzIxMTEwMjE5NjE5MjE5WjBbMQswCQYDVQQGEwJVUzETMBEGA1UECBMjQ2VydGlma
Zw1odG1sMDMwMjElBgNVBAMTBGFtcGxlLm5ldCAwggEICHQ8MB0GA1UdDwEB/wQEAwO
AKkBgkqhkiG9w0BAQgFAdAOAwIEAQDMMwGCCsGAQUFBwIwHjENTSAjBgNVHQ8BAf8EA
DcDADA0H0QHMA0GCSqGSIb3DQEBCwUAA4GKC6h+D6rZwX5U/vUhQvW75q6d0mQb7t3h
Wg40h3sokMIfp15c0/5eF86vZP75r5/oK0dM7s6t4gWpm8v2z9Y9396B97E55D2cyj
4Hc4/gV1t0wF0fVX8w1L7hr/sMyRZn7LQ7cX6h4lq6TB16FQsfa+6/NlV8wAvFZ5iz
2a24XnZqP2595DzG/wSxTQ9cRLBxgjnxHkOeCv9kS37p5d/vW/7+9R80+v+ps84Tb
-----END CERTIFICATE-----

Private key for domain certificate for example.net:
-----BEGIN RSA PRIVATE KEY-----
MIIEowIBAAKCAQEAgbHyDUpzc2FzQ5etuh3altHR+mrH2Xs7+0hYd1Z24AIuWq
0o8pKYrAVtna5/lT3k6reG5bEkbGcGb+1QL1LvMOGpWqJ2A8sxq+t30kVrN6H8
DAGN2T06HP/q2RReuvbL8XDDNoQgrD7xCbgufYTdts2pALio5edi3gsabLon17
riF9dWp7TN71x8opykFu3X9MXvm61ucfbbMQL+jVv7H9VIDOowoq034DvKvjW
sY1Z2V5UC4foa/XhHy3bTINQ1fjD7MWbd1JHO1aRvSiC7QRE17ueiCsK/2nfr8Z8i4
eksKwa00pJAGlzTvzgIPSFvmeW2ctWAn7s7GbwiDAQABaoIABAIjpv+B5YVITL59
+UCrd4JyKVLG1lioQf/CygafjztVTVa6a/ArNn8kgb8xYjr9sXvZVBlTqiUbdM4Z9I
8faVSKLAw3s7thkScjtMuz77x+1dCG6LxSzekAGqAI7sRL+iEz1/Fm1WlEyHl
GIW1LgH01n300eCy72dwmAV+2Ha/zn8eBggkWxMP0fb1RC9pVh0FCo+jyllHasjL
oc0Bh511bM4Pu4uy072j2z665mPn71On25igef842JkbqAV8Ao1lQ6y7t7RYEw
6QyLv0odebohrZ81EzahWAdm1PGC1UcFM7RmyInQatGAdVEU3uYnkQVo1/JT6
46CCmbEcgyEA4c1Dv/1Vz9pWlio/0MaJ94zfeg?Pgn5DRXnmNhJsCsSxVHS61W1UL
hCfyos77+WbTuixO2QeQe5m0A2ss3+4vKb80eue6EBQ/fE16CQQH52nxdV1sGk1n
5B5eiFkcZKPFnNVwr0BCcsDndTchp9STKsxWkEsL3C3V5zUXM5sosCGEyAEWNYV
+sScsCQcI782ZffkeYBnHqRUFKncx/tWQjop5qne4w53Lq13SNyCUqrdydE1wQDE3
6C0m197jc7JggDq7rgix1mxRXLMeG7bb7FwPE/SK0VH5uagEB7ktF18xIJK7
yOcK1ul1i0jloSs4uet6LRXXKCEFspRisW7wrDkCgYeA6DKBvYx/yikYDugp4d57
1pbFkDQ9G5x+YVTpUCx6g3gqafPjEANHhZVqoodTTRDTrYm7e8dpj22WiS3aB7WS
hfcTfCtveczcm+IDZ9GnKoVQ76aM6qC72j36eEBXUhPEa072Z8ZDCx1sWmEmeJN
+M2KhxcGIXt19eIeh3j1fowykCgyB9AUs1PwaeTVX130rdhuyUQ0xOoNmMA491Euh8
FpciP2DZtmzky2WvjeIPxFwQWlg1mMJZJeNeRPNpQrCRl65qzxKXJx/wBePn1PBm
tXLPrp6vnPkhJ+wnOeQ5ShkzKYvBv1hHa5icCuDx+pD4sWEmpmW+Gdn2FXCYwAF
UcKJ4QkBGaKsrr85yQOhd8RAMg9JZLGuPnmtTKNU98f3fUENxXZ7EZETasnn18vd
65x04h58choJkNnxel6k31c3MoPzZrsVsIha3ZMEoJFCgwBsa8zLzrrR13YQin6yf
+bAmFTDmhpOBRA36O4Y4B1kcwxKzQ0n3Aa1l17NRV5wHr2ezjK
-----END RSA PRIVATE KEY-----

B.3. Certificate Chaining with a Non-Root CA

Following is a certificate for a non-root CA in example.net. The certificate was signed by the root CA shown in Section 2.1. As indicated in Sections 4.2.1.9 and 4.2.1.3 [RFC5280], "cA" is set in Basic Constraints, and "keyCertSign" is set in Key Usage. This identifies the certificate holder as a signing authority.

Version: 3 (0x2)
Serial Number:
   96:a3:84:17:4e:ef:8a:52
Signature Algorithm: sha1WithRSAEncryption
Issuer: C=US, ST=California, L=San Jose, O=sipit,
   OU=Sipit Test Certificate Authority

Validity
   Not After: Jan 14 20:21:13 2111 GMT
Subject: C=US, ST=California, L=San Jose, O=sipit,
   OU=Test CA for example.net, CN=example.net
Subject Public Key Info:

Public Key Algorithm: rsaEncryption
RSA Public Key: (2048 bit)
  Modulus (2048 bit):
    dd:41:3f:7e:2a:e4:26:d5:a3:33:b0:5e:37:1d:ee:
    f9:0f:dc:9a:45:0f:28:8d:dd:fa:15:56:d5:35:17:
    50:ad:57:96:12:bb:87:1b:0c:01:03:bo:82:4d:0f:
    80:bf:dd:44:ff:ff:39:0a:2b:3e:4d:3f:5e:7a:aa:
    34:e5
Exponent: 65537 (0x10001)
X509v3 extensions:
  X509v3 Basic Constraints:
    CA:TRUE
X509v3 Subject Key Identifier:
X509v3 Authority Key Identifier:
X509v3 Key Usage:
  Certificate Sign
Signature Algorithm: sha1WithRSAEncryption
    fc:2:3e:00:ce:5f:eb:4c:8a:da:1b:5f:2f:cc:89:60:a4:3a:2b:be:
    61:62:6b:e5
Robert’s certificate was signed by the non-root CA in example.net:

Version: 3 (0x2)
Serial Number: 96:a3:84:17:4e:ef:8a:53
Signature Algorithm: sha1WithRSAEncryption
Issuer: C=US, ST=California, L=San Jose, O=sipit, OU=Test CA for example.net, CN=example.net
Validity
   Not After : Jan 14 20:21:13 2111 GMT
Subject: C=US, ST=California, L=San Jose, O=sipit, CN=robert
Subject Public Key Info:
   Public Key Algorithm: rsaEncryption
   RSA Public Key: (2048 bit)
      Modulus (2048 bit):
   Exponent: 65537 (0x10001)
X509v3 extensions:
   X509v3 Subject Alternative Name:
      URI:sip:robert@example.net, URI:im:robert@example.net, URI:pres:robert@example.net
   X509v3 Basic Constraints:
      CA:FALSE
   X509v3 Subject Key Identifier:
   X509v3 Authority Key Identifier:
X509v3 Key Usage:
Digital Signature, Non Repudiation, Key Encipherment

X509v3 Extended Key Usage:

- E-mail Protection, 1.3.6.1.5.5.7.3.20

Signature Algorithm: sha1WithRSAEncryption

Certificate for CA for example.net in PEM format:

-----BEGIN CERTIFICATE-----
MIIDzzCCAregAwIBAgIJAJajhBdO74pSMA0GCSqGSIb3DQEBBQUAMHAxCzAxBgNV
BAYTA1VTMRwEQEYDVQQIDAPYxpZm9ybmlhMREwDwYDVQQHDAHfTY4wSm9zZTFO
MAwGJ1UECgwFc2lwaXQzAQcCQg0GA1UdDwQEAwIBAQIBAd8wHQYDVR0OBBYEFH

Private key for CA for example.net:

Appendix C.  Message Dumps

This section contains a base64 encoded gzipped, compressed tar file of various Cryptographic Message Syntax (CMS) messages used in this document.  Saving the data in a file foo.tgz.b64 then running a command like "openssl base64 -d -in foo.tgz.b64 | tar xfz -" would recover the CMS messages and allow them to be used as test vectors.

```
-- BEGIN MESSAGE ARCHIVE --
H4sIAIpaUE0CA+ybeUATxx7HCSCIHiPozqISvFCu5tsDhAEDATQhCsQEXeT2Jb1yGUSIEREREUE8u12RqYERYRHUCqiKuBWPlvusXMlEeIvJ3lfpCaRuUpSF8f
6XJX/JRdmd3fJtzY/+f8r7J66ELKvvCw7yTCMd+yKhp/0LAAbEAgHb8EkI
wp98Whi5AQV centersA1CBcQrCtq1AG6v6EHKFWIQtsVrrKl59D9ZxVvt
jd8F+/hizcyl0+r+bgD5x0iU00hSHTRmndUjykYTRhttqwvb4BTpaAjYnoj
VIg4/37mxBwTgAuPIzINHyBFyBmEEAF24CKTk1L3LVkJoBO5yH9MgqkaHAIU
CwAqgSvAc39kwQDQgbzq19XyhlVumLNAcmGqADUeO82Y84RMFCXomN32hq
zBk7guv+uzN2BfJL+zCIQxKhh+saeATLomKtIxxKIWAsEHpMChpgYjH0
q+ymYu194gsdpw/py9nM33V2oAmWTDdpCQPQ6eFSgussopjYbVrIaj9rZg
dl7z1kw4bgC/vsFTHH48H1pxXWMLMNKVIU40eVE5KYaQU5TtgVhaha55QXIOKp
Axy8VCHQgisygQRO8mieRtmGOFJi1v5ccmZhC1VITU2xqBpMbGkKlapw
zhq5+sdnuXB1MV1ioco3JQjKAI0ULAhRaKMIINIBforUunjRVRiGiihWu814
MBrhw9D2p9EpopRoAM/FmwcKolUIWlfPsBN+2BwzW33rxhf79xkbbtOFK
UXo50S87Y8XS9GcsVyj9xVBPd84gsPsvsz/otz/TRYpIUNd8yfYI0Un58U6q
UcmXYYUN92qNgFQILQqirxeQqorR8bsynrgp0coxQx1/AqEybc3K78AKw15
b2A01sUn5YpxIaAMx1ZsfU1xXeDAYOa4N4U1830DpsDxh8wYDjmCztd9j
LWmAnlQLDQAyA0Vn7NVR7kGkQDzd9pU3rJO8UaAppaAKEdEq61Nqa2jSPRiR
Yvk15hS0QA0waqkx7MHTnjihCqrOHjvUGVjgtgxsxOxUqgKQRuQupatdAFy
jxV1eVhWVMxRumkZ8+0ZDyYS/1YQ9MPpCrugunoJ2N0VH7YoyoVx0Aazszi
RqsQ4fr+a1+26bndvntU7lf1KxyZUNqRPcseYUow/vk/O3E3C1l3ugTe5
5wu+dX+cWmm/2bj1U7N50VKeGmpk/cGM9cqt1D1U6qWM9qyg6l1/f03247
9cwy/tDpu53i8dePMXnyftLZsVBldeE0/9ke41EmjER19+4w54Kw67brci
Vd9g/bx6Ol1npn9Cq7MuVcTCCq+WG8137EnzmwFX0rZsCmsISuCpQja+J
1+5dXwd7/51lzHzntbT+f65yli/wN1rnp2q4e0c4F77exoxF0rH1xepGoyt
+2FPlkntYFq6m2j9osmpg+vv97jioXooaVBicO23t3EWCkPVJjIPt
izU4BaqU4dQ3eAip/AIEENOM/AH4ILv/HeU3pGmHbbZ/9xbzn/LjgP81Hokm
UPAEMFBvtw/JdKcp3pBYyglvehIgYlWSWpeErWUF01j0WKEWbQmTmSEAKH
GfiB910tS1+8rw7Hx97YDeiqFOjv/0r/0f9E9D+HhrVwN+nV/Vmdka15A
B6wOrsBra8X/-sG8U0+4BC1J79+XKJGTd3g7y/VCdxExOF6Hkg+e/8FsjQd
878z+P8rCohHfHfKSBMDAW21r+I/tqg/59FQeKD0oaTHY1gV1v/OF2Ce
PFFvjdpgEJQXUMAl8nU6MCSUhQZqiXW0nriU6M65x84R8E86CE8MSdD1dBCR
0L7+f3Pr7e1j/t2W6yf/8H8p+v9NPV5eaR67IYRFWfLfu+11h9PEH1Y17
3j7fs3bmkK0B0g3W2pqtgq9cboC6Od96zemFjQ4apsSHkbVgCoeTcTFK
OWy7cCOpAU/pugXq6mgQZPOdtBLNhvfVj+Vw/+utn/rowmayYapSrV8Dykr
neeNgsQnR8adKxop2bOKs3FLYdEzaeIqUusmMY5PsnVzdyFybytwua2/eLE
H1KrGu7dqepamG+9Wf8n880krFG0Zdhr1pCk2W383634htWhFLr3aen7xPzB
u4blweyC5s1Gz/Zt/LaHLkhZNaNd2YF9k604R0KkWaq1TOV5P0U5GtKAnVb
MxPU05e9HvBydpFDIe7tc1T/USkuD8V/8/A/2/kfyP7US1PES9xof2VgBt8R/C
```
Authors’ Addresses

Cullen Jennings
Cisco Systems
170 West Tasman Drive
Mailstop SJC-21/2
San Jose, CA  95134
USA

Phone: +1 408 421 9990
Email: fluffy@cisco.com

Kumiko Ono
Columbia University

Email: kumiko@cs.columbia.edu

Robert Sparks
Tekelec
17210 Campbell Road
Suite 250
Dallas, TX  75252
USA

Email: rjsparks@estacado.net

Brian Hibbard (editor)
Tekelec
17210 Campbell Road
Suite 250
Dallas, TX  75252
USA

Email: brian@estacado.net
Extending the Session Initiation Protocol (SIP) Reason Header for Applications
draft-mohali-sipcore-reason-extension-application-00

Abstract

This document defines a new protocol value "Application" for the Reason Header field and a new IANA Registry for registering application types. This new value enables signaling that a request or response has been issued as a result of a decision made by a particular type of application or that an initial request has been retargeted as a result of an application decision.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 16, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal
Table of Contents

1. Introduction .................................................. 3
2. Reason header extension .................................... 3
3. Use Cases ..................................................... 4
   3.1. Reason when issuing a Response ....................... 4
   3.2. Reason for Retargeting ................................ 5
4. IANA Considerations ........................................... 8
5. Security Considerations ....................................... 9
6. Acknowledgements ............................................. 9
7. Normative References ......................................... 9
Authors’ Addresses ............................................. 9
1. Introduction

The Reason header field [RFC3326] provides a mechanism to signal the reason why a SIP request or response was issued or why an initial request was retargeted. This document requests IANA registration of a new protocol value "Application" for the Reason Header field and a new IANA Registry for registering application types. This extension enables signaling that a request or response has been issued as a result of a decision made by a particular type of application or that an initial request has been retargeted as a result of an application decision (e.g. Freephone number translation).

2. Reason header extension

The syntax of the Reason Header as defined in [RFC3326] is as follows:

Reason            =  "Reason" HCOLON reason-value *(COMMA reason-value)
reason-value      =  protocol *(SEMI reason-params)
protocol          =  "SIP" / "Q.850" / token
reason-params     =  protocol-cause / reason-text
                    / reason-extension
protocol-cause    =  "cause" EQUAL cause
cause             =  1*DIGIT
reason-text       =  "text" EQUAL quoted-string
reason-extension  =  generic-param

This document adds a new protocol value "Application".

When this protocol value is used, the protocol-cause and the reason-text fields shall be set to values registered in a new IANA Registry for registering application types.

This document registers protocol causes for the following cases:

- Universal Access Number:
  A Universal Access Number application has influenced processing of the call (e.g. by translating a dialed Universal Access number to a routable directory number).

- Freephone:
  A Freephone application has influenced processing of the call (e.g. by translating a dialed Free Phone number to a routable
directory number).

- **Premium Rate:**
  A Premium Rate application has influenced processing of the call (e.g. by translating a dialed Premium Rate number to a routable directory number).

- **Directory Enquiry:**
  A Directory Enquiry application has influenced processing of the call (e.g. completion of the call to the Directory Number responding to the enquiry).

- **VPN:**
  A VPN application has influenced processing of the call (e.g. by translating a private number to a routable Directory Number).

- **Credit Card Calling:**
  A Credit Card Calling application has influenced processing of the call (e.g. it has completed the call to a directory number dialed in-band by the calling user).

- **Voice Activated Dialing:**
  A Voice-Activated Dialing application has influenced processing of the call (e.g. it has retargeted the call to a directory number determined by speech recognition).

- **Customized Call Routing:**
  An Customized Call Routing application has influenced processing of the call (e.g. by translating a dialed number into a routable directory number based on customized criteria such as time of day or caller’s location).

- **Prepaid:**
  A Prepaid application has influenced processing of the call (e.g. it has requested that the call be released due to exhaustion of credit available on the user account).

3. **Use Cases**

The following text provides two examples illustrating when this new class of Reason values could be used either as a result of a call completion failure or in a retargeting situation.

3.1. **Reason when issuing a Response**

This section presents the case where an application inserts the Reason header in a SIP response and wants to signal that a call is
A served user having a prepaid calling card wants to call a friend. First, the served user joins the Prepaid application where it gets authenticated. Then, the user dials a number corresponding to an international destination. Unfortunately, the user has not enough credits left on his card and at some point in time the Prepaid application ends the call by sending a BYE method with a Reason header set as follows:

Reason: application ;cause=9 ;text="Prepaid",
SIP ;cause=402; text="Payment Required"

Note: [RFC3326] states that a SIP message MAY contain more than one Reason value (i.e., multiple Reason lines), but all of them MUST have different protocol values and an implementation is free to ignore Reason values that it does not understand.

3.2. Reason for Retargeting

There are other situations where a call is retargeted as a result of an application decision (e.g. Freephone service). The proposed SIP extension of the Reason header allows to remain this application information in the signaling.

[RFC4244] defines a SIP header field, History-Info, to provide a standard mechanism for capturing requests history information. This allows receiving applications to determine hints about how and why the call arrived at the application/user. The History-Info header field can contain an optional reason, by including the Reason Header [RFC3326] escaped in the hi-targeted-to-uri field.

There are situations where a call is retargeted as a result of receiving a SIP response (e.g. 3XX response) in which case the Reason Header contains the SIP status code. There are other situations where a call is retargeted as a result of an application decision (e.g. Freephone service). The extension of the Reason header defined in the present document covers the second type of situation and allows to convey the identification of such applications in the signaling. This might help other applications invoked downstream to take appropriate decisions to avoid undesired service interactions.

The following example illustrates the case where the Reason header is set due to retargeting by a Freephone service running in a SIP Application Server.
Reason: application ;cause=2 ;text="Freephone"

More specifically, in such a case, the Reason header would appear in a History-Info header as described in [RFC4244] and illustrated below:

Alice calls the Company dialing a Freephone service number.
F1: INVITE Alice -> proxy.example.com

INVITE sip:+89955510044@example.com;user=phone SIP/2.0
Via: SIP/2.0/TCP proxy.example.com:5060;branch=z9hG4bK-klj79f
From: Alice <sip:alice@example.com>; tag=1234567
To: Company <sip:info@company.biloxie.com>
Supported: histinfo
Max-Forwards: 70
Call-Id: 12345600@example.com
Record-Route: <sip:proxy.example.com;lr>
CSeq: 1 INVITE
Contact: Alice <sip:alice@192.0.2.3>
Content-Type: application/sdp
Content-Length: <appropriate value>

<!--SDP Not Shown-->

F2: INVITE proxy.example.com -> Company

INVITE sip:+12125551212@phone2net.com SIP/2.0
Via: SIP/2.0/TCP as.example.com:5060;branch=z9hG4bK-vlk25e
Via: SIP/2.0/TCP proxy.example.com:5060;branch=z9hG4bK-klj79f
From: Alice <sip:alice@example.com>; tag=1234567
To: sip:+12125551212@phone2net.com;user=phone
Supported: histinfo
Max-Forwards: 69
Call-Id: 12345600@example.com
Record-Route: <sip:proxy.example.com;lr>
History-Info: <sip:+89955510044@example.com;user=phone?Reason:application
;cause=2;text="Freephone";index=1
<sip:+12125551212@phone2net.com>;index=1.1;mp=1
CSeq: 1 INVITE
Contact: Alice <sip:alice@192.0.2.3>
Content-Type: application/sdp
Content-Length: <appropriate value>

<!--SDP Not Shown-->
4. IANA Considerations

This document adds to one existing IANA Registry and creates one new Registry according to Section 27 of [RFC3261].

The existing IANA Registry for the SIP Reason Header is as follows:

<table>
<thead>
<tr>
<th>Protocol Value</th>
<th>Protocol Cause</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIP</td>
<td>Status code</td>
<td>RFC 3261</td>
</tr>
<tr>
<td>Q.850</td>
<td>Cause value in decimal</td>
<td>ITU-T Q.850</td>
</tr>
<tr>
<td>Preemption</td>
<td>Cause value in decimal*</td>
<td>RFC 4411</td>
</tr>
</tbody>
</table>

This document adds to that Registry with the following entry:

<table>
<thead>
<tr>
<th>Protocol Value</th>
<th>Protocol Cause</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td>Cause value in decimal</td>
<td>[RFCXXXX]</td>
</tr>
</tbody>
</table>

Below is a new IANA Registry for SIP Reason Header reason-text strings and reason-param cause values, associated with the protocol value "application". Per [RFC3326], the reason-text string is a quoted default string with only human understandability meant. These strings can be changed by local policy.

The following cause values and associated reason text are defined:

<table>
<thead>
<tr>
<th>Reason-</th>
<th>Protocol param</th>
<th>Reason-text</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Application</td>
<td>Universal Access Number</td>
<td>[RFCXXXX]</td>
</tr>
<tr>
<td></td>
<td>Application</td>
<td>Freephone</td>
<td>[RFCXXXX]</td>
</tr>
<tr>
<td></td>
<td>Application</td>
<td>Premium Rate</td>
<td>[RFCXXXX]</td>
</tr>
<tr>
<td></td>
<td>Application</td>
<td>Directory Enquiry</td>
<td>[RFCXXXX]</td>
</tr>
<tr>
<td></td>
<td>Application</td>
<td>VPN</td>
<td>[RFCXXXX]</td>
</tr>
<tr>
<td></td>
<td>Application</td>
<td>Credit Card Calling</td>
<td>[RFCXXXX]</td>
</tr>
<tr>
<td></td>
<td>Application</td>
<td>Voice Activated Dialing</td>
<td>[RFCXXXX]</td>
</tr>
<tr>
<td></td>
<td>Application</td>
<td>Customized Call Routing</td>
<td>[RFCXXXX]</td>
</tr>
<tr>
<td></td>
<td>Application</td>
<td>Prepaid</td>
<td>[RFCXXXX]</td>
</tr>
</tbody>
</table>

Note to the RFC-Editor: RFCXXXX should be replaced by this RFC reference when available.

The cause values created by the Application Protocol namespace in
this document are defined in Section 2. Each cause value has a Reason-text string as a general description of what the cause value is for. Section 3 provides an example for the above cause codes with a message flow diagram.

The above cause values are included in this registry. New cause values can be added and are allocated as First Come First Served.

5. Security Considerations

This specification does not add any security constraint beyond the security considerations for the Reason header field described in [RFC3326] apply.

6. Acknowledgements

7. Normative References


Authors' Addresses

Marianne Mohali
France Telecom Orange
38-40 rue du General Leclerc
Issy-Les-Moulineaux Cedex 9  92794
France

Phone: +33 1 45 29 45 14
Email: marianne.mohali@orange-ftgroup.com