Network Working Group
Request for Comments: 3262
Category: Standards Track
J. Rosenberg/H. Schulzrinne
dynamicsoft/Columbia U.
June 2002

Reliability of Provisional Responses in the Session Initiation Protocol (SIP)

Status of this Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Copyright Notice

Abstract

This document specifies an extension to the Session Initiation Protocol (SIP) providing reliable provisional response messages. This extension uses the option tag 100rel and defines the Provisional Response ACKnowledgement (PRACK) method.

Contents

1 Introduction

The Session Initiation Protocol (SIP) (RFC 3261 [1]) is a request-response protocol for initiating and managing communications sessions. SIP defines two types of responses, provisional and final. Final responses convey the result of the request processing, and are sent reliably. Provisional responses provide information on the progress of the request processing, but are not sent reliably in RFC 3261.

It was later observed that reliability was important in several cases, including interoperability scenarios with the PSTN. Therefore, an optional capability was needed to support reliable transmission of provisional responses. That capability is provided in this specification.

The reliability mechanism works by mirroring the current reliability mechanisms for 2xx final responses to INVITE. Those requests are transmitted periodically by the Transaction User (TU) until a separate transaction, ACK, is received that indicates reception of the 2xx by the UAC. The reliability for the 2xx responses to INVITE and ACK messages are end-to-end. In order to achieve reliability for provisional responses, we do nearly the same thing. Reliable provisional responses are retransmitted by the TU with an exponential backoff. Those retransmissions cease when a PRACK message is received. The PRACK request plays the same role as ACK, but for provisional responses. There is an important difference, however. PRACK is a normal SIP message, like BYE. As such, its own reliability is ensured hop-by-hop through each stateful proxy. Also like BYE, but unlike ACK, PRACK has its own response. If this were not the case, the PRACK message could not traverse proxy servers compliant to RFC 2543 [4].

Each provisional response is given a sequence number, carried in the RSeq header field in the response. The PRACK messages contain an RAck header field, which indicates the sequence number of the provisional response that is being acknowledged. The acknowledgments are not cumulative, and the specifications recommend a single outstanding provisional response at a time, for purposes of congestion control.

2 Terminology

In this document, the key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" are to be interpreted as described in RFC 2119 [2] and indicate requirement levels for compliant SIP implementations.

3 UAS Behavior

A UAS MAY send any non-100 provisional response to INVITE reliably, so long as the initial INVITE request (the request whose provisional response is being sent reliably) contained a Supported header field with the option tag 100rel. While this specification does not allow reliable provisional responses for any method but INVITE, extensions that define new methods that can establish dialogs may make use of the mechanism.

The UAS MUST send any non-100 provisional response reliably if the initial request contained a Require header field with the option tag 100rel. If the UAS is unwilling to do so, it MUST reject the initial request with a 420 (Bad Extension) and include an Unsupported header field containing the option tag 100rel.

A UAS MUST NOT attempt to send a 100 (Trying) response reliably. Only provisional responses numbered 101 to 199 may be sent reliably. If the request did not include either a Supported or Require header field indicating this feature, the UAS MUST NOT send the provisional response reliably.

> 100 (Trying) responses are hop-by-hop only. For this reason, the reliability mechanisms described here, which are end-to-end, cannot be used.

An element that can act as a proxy can also send reliable provisional responses. In this case, it acts as a UAS for purposes of that transaction. However, it MUST NOT attempt to do so for any request that contains a tag in the To field. That is, a proxy cannot generate reliable provisional responses to requests sent within the context of a dialog. Of course, unlike a UAS, when the proxy element receives a PRACK that does not match any outstanding reliable provisional response, the PRACK MUST be proxied.

There are several reasons why a UAS might want to send a reliable provisional response. One reason is if the INVITE transaction will take some time to generate a final response. As discussed in Section 13.3.1.1 of RFC 3261, the UAS will need to send periodic provisional responses to request an "extension" of the transaction at proxies. The requirement is that a proxy receive them every three minutes, but the UAS needs to send them more frequently (once a minute is recommended) because of the possibility of packet loss. As a more efficient alternative, the UAS can send the response reliably, in which case the UAS SHOULD send provisional responses once every two and a half minutes. Use of reliable provisional responses for extending transactions is RECOMMENDED.

The rest of this discussion assumes that the initial request contained a Supported or Require header field listing 100rel, and that there is a provisional response to be sent reliably.

The provisional response to be sent reliably is constructed by the UAS core according to the procedures of Section 8.2.6 of RFC 3261. In addition, it MUST contain a Require header field containing the option tag 100rel, and MUST include an RSeq header field. The value of the header field for the first reliable provisional response in a transaction MUST be between 1 and 2**31 - 1. It is RECOMMENDED that it be chosen uniformly in this range. The RSeq numbering space is within a single transaction. This means that provisional responses for different requests MAY use the same values for the RSeq number.

The reliable provisional response MAY contain a body. The usage of session descriptions is described in Section 5.

The reliable provisional response is passed to the transaction layer periodically with an interval that starts at T1 seconds and doubles for each retransmission (T1 is defined in Section 17 of RFC 3261). Once passed to the server transaction, it is added to an internal list of unacknowledged reliable provisional responses. The transaction layer will forward each retransmission passed from the UAS core.

> This differs from retransmissions of 2xx responses, whose intervals cap at T2 seconds. This is because retransmissions of ACK are triggered on receipt of a 2xx, but retransmissions of PRACK take place independently of reception of 1xx.

Retransmissions of the reliable provisional response cease when a matching PRACK is received by the UA core. PRACK is like any other request within a dialog, and the UAS core processes it according to the procedures of Sections 8.2 and 12.2.2 of RFC 3261. A matching PRACK is defined as one within the same dialog as the response, and whose method, CSeq-num, and response-num in the RAck header field match, respectively, the method from the CSeq, the sequence number from the CSeq, and the sequence number from the RSeq of the reliable provisional response.

If a PRACK request is received by the UA core that does not match any unacknowledged reliable provisional response, the UAS MUST respond to the PRACK with a 481 response. If the PRACK does match an unacknowledged reliable provisional response, it MUST be responded to with a 2xx response. The UAS can be certain at this point that the provisional response has been received in order. It SHOULD cease retransmissions of the reliable provisional response, and MUST remove it from the list of unacknowledged provisional responses.

If a reliable provisional response is retransmitted for 64*T1 seconds without reception of a corresponding PRACK, the UAS SHOULD reject the original request with a 5xx response.

If the PRACK contained a session description, it is processed as described in Section 5 of this document. If the PRACK instead contained any other type of body, the body is treated in the same way that body in an ACK would be treated.

After the first reliable provisional response for a request has been acknowledged, the UAS MAY send additional reliable provisional responses. The UAS MUST NOT send a second reliable provisional response until the first is acknowledged. After the first, it is RECOMMENDED that the UAS not send an additional reliable provisional response until the previous is acknowledged. The first reliable provisional response receives special treatment because it conveys the initial sequence number. If additional reliable provisional responses were sent before the first was acknowledged, the UAS could not be certain these were received in order.

The value of the RSeq in each subsequent reliable provisional response for the same request MUST be greater by exactly one. RSeq numbers MUST NOT wrap around. Because the initial one is chosen to be less than 2**31 - 1, but the maximum is 2**32 - 1, there can be up to 2**31 reliable provisional responses per request, which is more than sufficient.

The UAS MAY send a final response to the initial request before having received PRACKs for all unacknowledged reliable provisional responses, unless the final response is 2xx and any of the unacknowledged reliable provisional responses contained a session description. In that case, it MUST NOT send a final response until those provisional responses are acknowledged. If the UAS does send a final response when reliable responses are still unacknowledged, it SHOULD NOT continue to retransmit the unacknowledged reliable provisional responses, but it MUST be prepared to process PRACK requests for those outstanding responses. A UAS MUST NOT send new reliable provisional responses (as opposed to retransmissions of unacknowledged ones) after sending a final response to a request.

4 UAC Behavior

When the UAC creates a new request, it can insist on reliable delivery of provisional responses for that request. To do that, it inserts a Require header field with the option tag 100rel into the request. A Require header with the value 100rel MUST NOT be present in any requests excepting INVITE, although extensions to SIP may allow its usage with other request methods.

If the UAC does not wish to insist on usage of reliable provisional responses, but merely indicate that it supports them if the UAS needs to send one, a Supported header MUST be included in the request with the option tag 100rel. The UAC SHOULD include this in all INVITE requests.

If a provisional response is received for an initial request, and that response contains a Require header field containing the option tag 100rel, the response is to be sent reliably. If the response is a 100 (Trying) (as opposed to 101 to 199), this option tag MUST be ignored, and the procedures below MUST NOT be used.

The provisional response MUST establish a dialog if one is not yet created.

Assuming the response is to be transmitted reliably, the UAC MUST create a new request with method PRACK. This request is sent within the dialog associated with the provisional response (indeed, the provisional response may have created the dialog). PRACK requests MAY contain bodies, which are interpreted according to their type and disposition.

Note that the PRACK is like any other non-INVITE request within a dialog. In particular, a UAC SHOULD NOT retransmit the PRACK request when it receives a retransmission of the provisional response being acknowledged, although doing so does not create a protocol error.

Once a reliable provisional response is received, retransmissions of that response MUST be discarded. A response is a retransmission when its dialog ID, CSeq, and RSeq match the original response. The UAC MUST maintain a sequence number that indicates the most recently received in-order reliable provisional response for the initial request. This sequence number MUST be maintained until a final response is received for the initial request. Its value MUST be initialized to the RSeq header field in the first reliable provisional response received for the initial request.

Handling of subsequent reliable provisional responses for the same initial request follows the same rules as above, with the following difference: reliable provisional responses are guaranteed to be in order. As a result, if the UAC receives another reliable provisional response to the same request, and its RSeq value is not one higher than the value of the sequence number, that response MUST NOT be acknowledged with a PRACK, and MUST NOT be processed further by the UAC. An implementation MAY discard the response, or MAY cache the response in the hopes of receiving the missing responses.

The UAC MAY acknowledge reliable provisional responses received after the final response or MAY discard them.

5 The Offer/Answer Model and PRACK

RFC 3261 describes guidelines for the sets of messages in which offers and answers [3] can appear. Based on those guidelines, this extension provides additional opportunities for offer/answer exchanges.

If the INVITE contained an offer, the UAS MAY generate an answer in a reliable provisional response (assuming these are supported by the UAC). That results in the establishment of the session before completion

Header field	where	PRACK
Accept	R	o
Accept	2xx	-
Accept	415	c
Accept-Encoding	R	o
Accept-Encoding	2xx	-
Accept-Encoding	415	c
Accept-Language	R	o
Accept-Language	2xx	-
Accept-Language	415	c
Alert-Info	R	-
Alert-Info	180	-
Allow	R	o
Allow	2xx	o
Allow	r	o
Allow	405	m
Authentication-Info	2xx	o
Authorization	R	o
Call-ID	c	m
Call-Info		-
Contact	R	-
Contact	1xx	-
Contact	2xx	-
Contact	3xx	o
Contact	485	o
Content-Disposition		o
Content-Encoding		o
Content-Language		o
Content-Length		t
Content-Type		*
CSeq	c	m
Date		o
Error-Info	300-699	o
Expires		-
From	c	m
In-Reply-To	R	-
Max-Forwards	R	m
Min-Expires	423	-
MIME-Version		o
Organization		-

Table 1: Summary of header fields, A–O

Header field	where	PRACK
Priority	R	-
Proxy-Authenticate	407	m
Proxy-Authenticate	401	o
Proxy-Authorization	R	o
Proxy-Require	R	o
Record-Route	R	o
Record-Route	2xx,18x	o
Reply-To		-
Require		c
Retry-After	404,413,480,486	o
	500,503	o
	600,603	o
Route	R	c
Server	r	o
Subject	R	-
Supported	R	o
Supported	2xx	o
Timestamp		o
To	c	m
Unsupported	420	m
User-Agent		o
Via	c	m
Warning	r	o
WWW-Authenticate	401	m

Table 2: Summary of header fields, P–Z

of the call. Similarly, if a reliable provisional response is the first reliable message sent back to the UAC, and the INVITE did not contain an offer, one MUST appear in that reliable provisional response.

If the UAC receives a reliable provisional response with an offer (this would occur if the UAC sent an INVITE without an offer, in which case the first reliable provisional response will contain the offer), it MUST generate an answer in the PRACK. If the UAC receives a reliable provisional response with an answer, it MAY generate an additional offer in the PRACK. If the UAS receives a PRACK with an offer, it MUST place the answer in the 2xx to the PRACK.

Once an answer has been sent or received, the UA SHOULD establish the session based on the parameters of the offer and answer, even if the original INVITE itself has not been responded to.

If the UAS had placed a session description in any reliable provisional response that is unacknowledged when the INVITE is accepted, the UAS MUST delay sending the 2xx until the provisional response is acknowledged. Otherwise, the reliability of the 1xx cannot be guaranteed, and reliability is needed for proper operation of the offer/answer exchange.

All user agents that support this extension MUST support all offer/answer exchanges that are possible based on the rules in Section 13.2 of RFC 3261, based on the existence of INVITE and PRACK as requests, and

2xx and reliable 1xx as non-failure reliable responses.

6 Definition of the PRACK Method

This specification defines a new SIP method, PRACK. The semantics of this method are described above. Tables 1 and 2 extend Tables 2 and 3 from RFC 3261 for this new method.

7 Header Field Definitions

This specification defines two new header fields, RAck and RSeq. Table 1 extends Tables 2 and 3 from RFC 3261 for these headers.

7.1 RSeq

The RSeq header is used in provisional responses in order to transmit them reliably. It contains a single numeric value from 1 to 2**32 – 1. For details on its usage, see Section 3.

Example:

```
RSeq: 988789
```

Header field	where	proxy	ACK	BYE	CAN	INV	OPT	REG	PRA
RAck	R		-	-	-	-	-	-	m
RSeq	1xx		-	-	-	o	-	-	-

Figure 1: RAck and RSeq Header Fields

7.2 RAck

The RAck header is sent in a PRACK request to support reliability of provisional responses. It contains two numbers and a method tag. The first number is the value from the RSeq header in the provisional response that is being acknowledged. The next number, and the method, are copied from the CSeq in the response that is being acknowledged. The method name in the RAck header is case sensitive.

Example:

```
RAck: 776656 1 {\INVITE}
```

8 IANA Considerations

This document registers a new option tag and two new headers, based on the IANA registration process of RFC 3261.

8.1 IANA Registration of the 100rel Option Tag

This specification registers a single option tag, 100rel. The required information for this registration, as specified in RFC 3261, is:

Name: 100rel

Description: This option tag is for reliability of provisional responses. When present in a Supported header, it indicates that the UA can send or receive reliable provisional responses. When present in a Require header in a request, it indicates that the UAS MUST send all provisional responses reliably. When present in a Require header in a reliable provisional response, it indicates that the response is to be sent reliably.

8.2 IANA Registration of RSeq and RAck Headers

The following is the registration for the RSeq header:

RFC Number: RFC3262

Header Name: RSeq

Compact Form: none

The following is the registration for the RAck header:

RFC Number: RFC3262

Header Name: RAck

Compact Form: none

9 Security Considerations

The PRACK request can be injected by attackers to force retransmissions of reliable provisional responses to cease. As these responses can convey important information, PRACK messages SHOULD be authenticated as any other request. Authentication procedures are specified in RFC 3261.

10 Collected BNF

The BNF for the RAck and RSeq headers and the PRACK method are defined here.

```
PRACKm        =  %x50.52.41.43.4B ; PRACK in caps
Method        =  INVITEm / ACKm / OPTIONSm / BYEm
                 / CANCELm / REGISTERm / PRACKm
```

```
                    / extension-method
RAck            =   "RAck" HCOLON response-num LWS CSeq-num LWS Method
response-num    =   1*DIGIT
CSeq-num        =   1*DIGIT
RSeq            =   "RSeq" HCOLON response-num
```

11 Acknowledgements

The authors would like to thank Jo Hornsby, Jonathan Lennox, Rohan Mahy, Allison Mankin, Adam Roach, and Tim Schroeder for the comments on this document.

12 Authors' Addresses

Jonathan Rosenberg
dynamicsoft
72 Eagle Rock Avenue
First Floor
East Hanover, NJ 07936
EMail: jdrosen@dynamicsoft.com

Henning Schulzrinne
Columbia University
M/S 0401
1214 Amsterdam Ave.
New York, NY 10027-7003
EMail: schulzrinne@cs.columbia.edu

Full Copyright Statement

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.

Normative References

[1] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, “SIP: session initiation protocol,” RFC 3261, Internet Engineering Task Force, June 2002.

[2] S. Bradner, “Key words for use in RFCs to indicate requirement levels,” RFC 2119, Internet Engineering Task Force, Mar. 1997.

[3] J. Rosenberg and H. Schulzrinne, “An offer/answer model with session description protocol (SDP),” RFC 3264, Internet Engineering Task Force, June 2002.

Informative References

[4] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg, “SIP: session initiation protocol,” RFC 2543, Internet Engineering Task Force, Mar. 1999.