The SIP PING Method
draft-fwmiller-ping-03

The Session Initiation Protocol (SIP) has the potential for long periods of time to elapse when no signaling traffic is sent between a User Agent Client (UAC) and a User Agent Server (UAS). There are situations when it may be necessary for some signaling traffic to flow periodically between these endpoints or to have a quick, lightweight check for whether a UAS is alive. The PING method is proposed that can be used for these purposes.
Table of Contents

1. Introduction .................................................. 3
2. PING Method .................................................... 4
  2.1. Header Field Support for PING Method .................... 4
  2.2. Response to the PING Method ............................... 6
  2.3. Message Body Inclusion ................................. 6
  2.4. User Agent Behavior .................................. 6
  2.5. Behavior of SIP Proxy and Redirect Servers ............ 7
    2.5.1. Proxy Server .................................. 7
    2.5.2. Forking Proxy Server ............................. 7
    2.5.3. Redirection Server ............................... 7
3. Security Considerations ................................. 8
4. References .................................................. 8

Author’s Address ........................................ 9
Intellectual Property and Copyright Statements ................. 10
1. Introduction

Two SIP entities pass signaling traffic between them as required to support SIP-based services. There can be long periods of time either when session is established or when no session exists when no traffic is flowing between the endpoints. There are situations where some signaling traffic should be sent during these long intervals between the UAC and UAS. For example, if one of the endpoints is behind a Network Address Translation (NAT), signaling traffic may be used to keep the NAT port bindings alive. The PING method is intended to confirm that the endpoints are alive and verify that a signaling path is still valid.
2. PING Method

The PING method is used to determine if a UAS is alive and to validate a signaling path. The PING method is not used to change the state of SIP calls, nor does it change the state of sessions initiated by SIP. Rather, it provides an indication to both ends of a session that signaling messages can still flow between them.

A PING request may be sent at any time. PINGS may be sent periodically to serve as a heartbeat. A UAC MUST NOT have more than one outstanding PING transaction in existence at any time with a specific UAS. A UAC MUST space PING transactions with the same UAS at least 500 milliseconds apart.

A PING request is routed the same way any other request is routed. This can be either direct signaling between the UAC and UAS or a signaling path involving SIP servers that potentially add themselves to the Record-Route headers. The sending of a PING request initiates a nominal non-INVITE transaction as specified in Section 17.1.2 of [1].

2.1. Header Field Support for PING Method

The PING request does not carry any information other than the intent to check for the liveness of the UAS and the signaling path validity. As such, only a few headers are used in both the PING request and its associated response.

<table>
<thead>
<tr>
<th>Header</th>
<th>Where</th>
<th>PING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accept</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Accept-Encoding</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Accept-Language</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Alert-Info</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Allow</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Authentication-Info</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Authorization</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Call-ID</td>
<td>R</td>
<td>m</td>
</tr>
<tr>
<td>Call-ID</td>
<td>200</td>
<td>m</td>
</tr>
<tr>
<td>Call-Info</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Contact</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Content-Disposition</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Content-Encoding</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Content-Language</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Content-Length</td>
<td>R</td>
<td>t</td>
</tr>
<tr>
<td>Content-Type</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CSeq</td>
<td>R</td>
<td>m</td>
</tr>
<tr>
<td>CSeq</td>
<td>200</td>
<td>m</td>
</tr>
</tbody>
</table>
If a Content-Length field is included in the PING request, it MUST be set to zero (0).

The intent is to provide as simple a message as possible to allow for implementations (particularly for servers) to optimize PING message processing.
2.2. Response to the PING Method

A UAC formats a PING request as desired and sends it to the UAS. The request is sent using the same retransmission and routing rules by which an OPTIONS non-INVITE transaction would be sent.

\[
\begin{array}{c|c}
\text{UAC} & \text{UAS} \\
\hline
\text{PING} & \text{200 OK} \\
\text{--------->} & <----------}
\end{array}
\]

There is only one defined response to a PING message. This means that a UAS that receives, recognizes, and supports the PING method MUST only send one possible response back to the UAC.

The defined response is a 200 OK response. A UAS that supports reception of the PING method MUST respond IMMEDIATELY with a 200 OK message when it receives a PING request.

If a UAS that does not support the PING method receives a PING request, it will generate other responses, e.g. a 501 Not Implemented per [1]. A UAC SHOULD accept any response other than a 1xx provisional response or a 3xx redirection. If a response other than a 1xx or a 3xx is received, the UAC SHOULD assume that the UAS does not recognize or support the PING method but the UAC SHOULD accept the response as if it were a 200 OK response. A UAC receiving a 1xx or 3xx response SHOULD drop the response as if it were never received.

2.3. Message Body Inclusion

A PING request MUST NOT contain a message body.

2.4. User Agent Behavior

Unless otherwise stated, the protocol rules for the PING request governing the usage of tags, Route, and Record-Route, retransmission and reliability, CSeq incrementing and message formatting follow those in [1] as defined for the OPTIONS request.

An implementation may want to optimize the processing of received PING requests. One potential implementation optimization is to recognize the PING method on the request line by scanning the first four characters of an incoming request for the PING method name.
Implementations may then initiate expedited processing of the request in order to return a 200 OK response as quickly as possible. For example, an implementation may simply copy the To, From, Via, Call-ID, and CSeq headers from the request into the response.

A PING request MAY NOT be canceled.

2.5. Behavior of SIP Proxy and Redirect Servers

2.5.1. Proxy Server

Unless stated otherwise, the protocol rules for the PING request at a proxy are identical to those for a OPTIONS request as specified in [1].

2.5.2. Forking Proxy Server

Unless stated otherwise, the protocol rules for the PING request at a proxy are identical to those for a OPTIONS request as specified in [1].

2.5.3. Redirection Server

Unless stated otherwise, the protocol rules for the PING request at a proxy are identical to those for a OPTIONS request as specified in [1].
3. Security Considerations

The only security consideration is that of a Denial of Service (DoS). A "PING Storm" DoS attack can be launched at a UAS if PING requests are sent at closer intervals than 500 milliseconds. Even 500 milliseconds can be considered tight. It is RECOMMENDED that PING request intervals be at least several seconds if possible.

4. References

Author’s Address

Frank W. Miller
Cornfed Systems, LLC
103 Overhill Road
Baltimore, MD  21210
US

Phone: +1 410 404 8790
Email: fwmiller@cornfed.com
URI:  http://www.cornfed.com/
Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Copyright Statement

Copyright (C) The Internet Society (2006). This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

Acknowledgment

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