SIP Performance Metrics
draft-malas-performance-metrics-01.txt

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of 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 November 25, 2006.

Abstract

This document defines the use of industry recommended reliability metrics for use with the SIP.

Conventions 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 RFC-2119 [1].

Table of Contents

1. Introduction...................................................2
1. Introduction

SIP has become a standard among many service providers, vendors, and end users. Although there are many different standards for measuring the performance of signaling protocols, none of these have been adapted for use with SIP. This document is intended for providing a guideline for the above listed entities in providing a standard approach for measuring and reporting SIP performance metrics in a production environment. This will allow a common approach and understanding of expectations between service providers, vendors, and the users of those services.

Not all metrics for performance map to all applications of the SIP. This document provides an overview of many different metrics, which may be used as an individual or set of metrics necessary based on the use of SIP.

There are many metrics available for determining performance. Although this document contains a number of them, it is not intended to be exhaustive. Instead, it is designed to provide a common sub-set with a common agreed upon definition.

Although these metrics may be used in a test environment, the IETF Benchmarking Methodology working group is currently working on a draft for this purpose. This draft will align terminology and methodologies where applicable, in order to maintain consistency among the IETF.
2. SIP Performance Metrics

The following metrics may be utilized for all applications.

2.1. Session SetupDelay (SSD)

In a successful request attempt, SSD is defined as the time interval from the moment the INVITE message containing the necessary information is passed by the originating agent or user to the intended mediation or destination agent until a response is received indicating an audible or visual status of the INVITE request. In SIP, the message indicating status would be a 180 and 183 message received in response to an INVITE request. In some cases, a 180 or 183 message is not received, but rather a 200 message is received as the first status message instead. In these situations, the 200 message would be used to calculate the interval.

In a failed request attempt, the interval is defined from the INVITE request and a failure indication status response. A failure response is described as a 4XX, 5XX, or possible 6XX message.

\[
SSD = \text{Time of Status Indicative Response} - \text{Time of INVITE}
\]

\[
\text{ASSD} = \frac{\text{SUM (Time of Status Indicative Response - Time of INVITE)}}{\text{SUM # of INVITE Requests}}
\]

ASSD = Average SSD

The following flow provides an example of Session Setup Delay:

```
UA1                      UA2
INVITE                   |
------------------------>
         \ /  100
<---------||-------->
         SSD
         \|
         \ /  180
<---------------
```

Malas                  Expires November 25, 2006
2.2. Session Completion Delay (SCD)

SCD is defined as the interval between sending a session completion message, such as a BYE, and receiving the subsequent 2XX acknowledgement. The following flow provides an example of this metric:

```
UA1                     UA2
|                       |
| INVITE                 |------------------------>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
</tr>
<tr>
<td>&lt;----------------------</td>
</tr>
<tr>
<td>180/200</td>
</tr>
<tr>
<td>&lt;----------------------</td>
</tr>
<tr>
<td>BYE</td>
</tr>
<tr>
<td>&lt;----------------------</td>
</tr>
<tr>
<td>|</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>SCD</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>200 \</td>
</tr>
</tbody>
</table>
|------------------------>
```

This calculation can also be extended to include teardown of specific legs of a call with the use of extensions such as REFER or capabilities such as RE-INVITE’s.

2.3. Session Duration Time (SDT)

SDT is usually calculated as an average and is defined as the duration of a call from receipt of a 200 OK and an associated BYE message indicating call completion.

\[
SDT = \text{Time of BYE} - \text{Time of 200 OK}
\]

\[
\text{ASDT} = \frac{\sum (\text{Time of BYE} - \text{Time of 200 OK})}{\sum \# \text{ of INVITE w/ 200OK & BYE}}
\]

ASDT = Average SDT

The following flow represents an example of the determination of this metric:
2.4. Sessions Per Second (SPS)

SPS is described as the number of calls, which are setup in an incremental time period of one second. In order for a call to be setup, an INVITE must be processed with a subsequent acknowledgement response from the terminating UA or UAS associated with the initial INVITE. The following flow provides an example of a "session" related to this metric:

2.5. Session Attempts Per Second (SAPS)

SAPS is defined as the number of initial INVITE requests received by a UA or UAS per time increment of one second. Usually, this metric is relative to proxy servers and the maximum number of SAPS it is capable of processing before failure conditions begin to occur. In addition, the metric may be used to trend a specific traffic pattern. This metric may be extended to include any initial SIP related
requests in addition to INVITE’s, such as REGISTER requests, SUBSCRIBE requests, and others as described in [2] and related extensions. The following flow provides an example of a call attempt:

```
UA1                     UA2
|                       |
| INVITE                 |
| ---------------------> |
```

As is described by the figure, a session attempt does not require a response from UA2 in order to constitute an attempt; however, an attempt may be acknowledged in order to validate the terminating UA or UAS received the attempt.

2.6. Session Establishment Rate (SER)

SER is defined as the number of INVITE requests resulting in a 200 OK response, to the total number of attempted INVITE requests.

\[
\text{SER} = \frac{\text{# of INVITE Requests with associated 200OK}}{\text{Total # of INVITE Requests}}
\]

The following flow represents session establishment as described above:

```
UA1                     UA2
|                       |
| INVITE                 |
| ---------------------> |
| 100                   |
| <---------------------|
| Session Established   |
| 180                   |
| <---------------------|
| 200                   |
```

In the event of re-transmissions of an INVITE, the SER calculation should include only one INVITE in the initial setup of a session dialog.
3. Security Considerations

Security should be considered in the aspect of securing the relative data utilized in providing input to the above calculations. All other aspects of security should be considered as described in [2].

4. IANA Considerations

There are no IANA considerations at this time.

5. Conclusions

The proposed guideline provides a description of common performance metrics, and their defined use with SIP. The use of these metrics will provide a common viewpoint across all vendors, service providers, and customers. These metrics will likely be utilized in production SIP environments for providing input regarding Key Performance Indicators (KPI) and Service Level Agreement (SLA) indications.

6. Acknowledgments

TBD

7. References

7.1. Normative References


7.2. Informative References

Author’s Addresses

Daryl Malas  
Level 3 Communications LLC  
1025 Eldorado Blvd.  
Broomfield, CO 80021  
USA  
EMail: daryl.malas@level3.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.