

**SIP End-to-End Performance Metrics**  
**draft-malas-performance-metrics-08.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 May 6, 2008.

Abstract

This document defines a set of metrics and their usage to evaluate the performance of end-to-end Session Initiation Protocol (SIP) based services in both production and testing environments. The purpose of this document is to combine a set of common metrics, allowing interoperable performance measurements, easing the comparison of industry implementations.

Table of Contents

<a href="#">1.</a>	<a href="#">Introduction.....</a>	<a href="#">2</a>
<a href="#">2.</a>	<a href="#">Terminology.....</a>	<a href="#">3</a>
<a href="#">3.</a>	<a href="#">SIP Performance Metrics.....</a>	<a href="#">4</a>
<a href="#">3.1.</a>	<a href="#">Registration Request Delay (RRD).....</a>	<a href="#">5</a>
<a href="#">3.1.1.</a>	<a href="#">Successful REGISTER Completion RRD.....</a>	<a href="#">5</a>
<a href="#">3.1.2.</a>	<a href="#">Failed REGISTER Attempt RRD.....</a>	<a href="#">6</a>

Malas

Expires June 6, 2008

[Page 1]

3.2.	Session Request Delay (SRD).....	7
3.2.1.	Successful Session Setup SRD.....	7
3.2.2.	Failed Session Setup SRD.....	8
3.2.3.	Instant Messaging.....	9
3.3.	Session Disconnect Delay (SDD).....	9
3.3.1.	Successful session completion SDD.....	10
3.3.2.	Failed session completion SDD.....	11
3.4.	Session Duration Time (SDT).....	12
3.4.1.	Successful session completion SDT.....	13
3.4.2.	Failed session completion SDT.....	14
3.5.	Average Hops per Request (AHR).....	15
3.6.	Session Establishment Rate (SER).....	17
3.6.1.	Instant Messaging.....	18
3.7.	Session Establishment Efficiency Rate (SEER).....	18
3.8.	Session Defects (SD).....	19
3.9.	Ineffective Session Attempts (ISA).....	19
3.10.	Session Disconnect Failures (SDF).....	20
3.11.	Session Completion Rate (SCR).....	20
3.11.1.	Successful Session Completion.....	21
3.11.2.	Failed Session Completion.....	22
3.12.	Session Success Rate (SSR).....	22
4.	Metric Correlations.....	23
5.	Additional Considerations.....	23
5.1.	Back-to-back User Agent (B2BUA).....	23
5.2.	Authorization and Authentication.....	23
5.3.	Forking.....	24
5.4.	Data Collection.....	24
5.5.	Testing Documentation.....	24
5.6.	Metric Template.....	25
6.	Security Considerations.....	25
7.	IANA Considerations.....	25
8.	Conclusions.....	25
9.	Contributors.....	26
10.	Acknowledgments.....	26
11.	References.....	26
11.1.	Normative References.....	26
11.2.	Informative References.....	27
	Author's Addresses.....	27
	Intellectual Property Statement.....	27
	Disclaimer of Validity.....	28
	Copyright Statement.....	28
	Acknowledgment.....	28

## **1. Introduction**

SIP has become a widely-used standard among many service providers, vendors, and end users. Although there are many different standards for measuring the performance of signaling protocols, none of them

specifically address SIP.

Malas

Expires June 6, 2008

[Page 2]

The scope of this document is limited to the definitions of a standard set of metrics for measuring and reporting SIP performance from an end-to-end perspective. The metrics introduce a common foundation for understanding and quantifying performance expectations between service providers, vendors, and the users of services based on SIP.

Measurements of the metrics described in this document are affected by variables external to SIP. The following is a non-exhaustive list of examples:

- . Network connectivity
- . Switch and router performance
- . Server processes and hardware performance

Note that some metrics in this document may not apply to all applications of SIP. This document provides an overview of pertinent metrics, which may be used individually or as a set based on the usage of SIP within the context of a given service.

The metrics defined in this document DO NOT take into consideration the impairment or failure of actual application processing of a request or response. The metrics do not distinguish application processing time from other sources of delay, such as packet transfer delay.

Metrics designed to quantify single device application processing performance are beyond the scope of this document.

This document does not provide any numerical objectives or acceptance threshold values for the SIP performance metrics defined below, as these items are beyond the scope of IETF activities, in general.

## **2. Terminology**

The following terms and conventions will be used throughout 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].

End-to-End - This is described as two or more elements utilized for initiating a request, receiving the request, and responding to the request. It encompasses elements as necessary to be involved in a session dialog between the originating user agent client (UAC), destination user agent server (UAS), and any interim proxies (may

Malas

Expires June 6, 2008

[Page 3]

also include back-to-back user agent's (B2BUA's)). This may be relative to a single operator's set of elements or extend to encompass all elements (if beyond a single operator's network) associated with a session.

Session - As described in [RFC 3261](#), SIP is used primarily to request, create, and conclude sessions. "These sessions include Internet telephone calls, multimedia distribution, and multimedia conferences [2]." The metrics within this document measure the performance associated with the processes necessary to establish these sessions; therefore, they are titled as: Session Request Delay, Session Disconnect Delay, etc. Although the titles of many of the metrics include this term, they are specifically measuring the signaling aspects only.

Session Establishment - Session establishment occurs when a 200 OK response from the UAS has been received, in response to a corresponding UAC's INVITE setup request, indicating the session setup request was successful.

Session Setup - As referenced within the sub-sections of 3.2 in this document, session setup is the set of messages and included parameters directly related to the process of a UAC requesting to establish a session with a corresponding UAS. This is also described as a set of steps in order to establish "ringing" [2].

Time Begin (TB) - This is the time instant that starts a continuous time interval running when a request is sent. TB occurs when the designated request has been processed by the SIP application and the first bit of the request packet has been sent from the proxy or UA (and is externally observable at some logical or physical interface).

Time Stop (TS) - This is the time instant that ends a continuous time interval running from when the related request is sent. TS occurs when the last bit of the designated response is received by the SIP application at the requesting device (and is externally observable at some logical or physical interface).

### **3. SIP Performance Metrics**

In regards to all of the following metrics, TB begins with the first associated SIP message sent by the UAC or UAS, and is not reset if the UAC or UAS must retransmit the same request multiple times. The first associated SIP message indicates the TB associated with the user or application expectation relative to the request.

Some metrics are calculated based on the final response message. These metrics do not take into consideration route advances to additional signaling functions based on "final" failure responses.

Malas

Expires June 6, 2008

[Page 4]

In these unique cases, the final response related to the initial setup attempt should be utilized for input to the metric.

In regards to all of the metrics, the output values are directly related to the accuracy and the equivalent level of granularity of the input values.

The following metrics may be utilized for many different SIP applications.

### **3.1. Registration Request Delay (RRD)**

Registration Request Delay is utilized to detect failures or impairments causing delays in responding to a UAC REGISTER request. RRD is measured for both successful and failed REGISTER requests. This metric is measured at the UAC only. The output value of this metric is numerical and should be adjusted to indicate milliseconds. The following represents the calculation for this metric:

$$\text{RRD} = \text{Time of Final Response} - \text{Time of REGISTER Request}$$

An average is one of the uses of this metric. The following represents the calculation for this metric as an average:

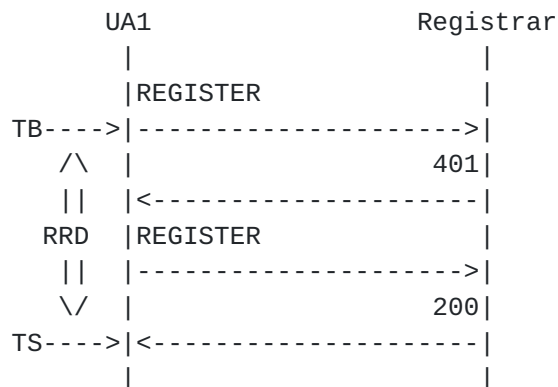
$$\text{ARRD} = \frac{\text{SUM (Time of Final Response - Time of REGISTER Request)}}{\text{SUM \# of REGISTER Requests}}$$

#### **3.1.1. Successful REGISTER Completion RRD**

In a successful registration attempt, RRD is defined as the time interval from the moment the initial REGISTER message containing the necessary information is passed by the originating UAC to the intended registrar until the 200OK is received indicating the registration attempt has completed successfully. This dialog includes an expected authentication challenge prior to receiving the 200OK as describe in the following registration flow examples.

The following flow provides an example of identifiable events necessary for inputs in calculating RRD during a successful registration completion:

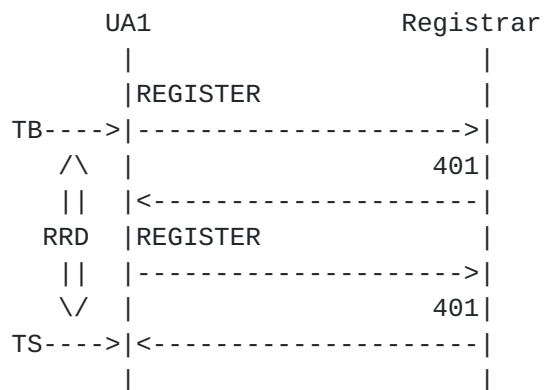




### 3.1.2. Failed REGISTER Attempt RRD

In a failed registration attempt, the interval is defined from the initial REGISTER request and the final response indicating a failure received from the destination registrar or interim proxies (Note: Many times registration attempts are repeated; therefore, a failed scenario must identify a failure response associated with the final attempt). A failure response is described as a 4XX, 5XX, or possible 6XX message. RRD may be used to detect problems in downstream signaling functions, which may be impairing the REGISTER message from reaching the intended registrar.

The following flow provides an example of identifiable events necessary for inputs in calculating RRD during a failed registration attempt:



Note: A second 401 was used due to a common failure related to incorrect authentication credentials related to the second REGISTER request after the initial request failed. In this case, the UAC gives up on attempting to REGISTER with the registrar.

Malas

Expires June 6, 2008

[Page 6]

### 3.2. Session Request Delay (SRD)

Session Request Delay is utilized to detect failures or impairments causing delays in responding to a UA session request. SRD is measured for both successful and failed session setup requests. This metric is similar to Post-Selection Delay [9] or Post-Dial Delay (PDD) in telephony applications of SIP, and it is measured at the UAC only. The output value of this metric is numerical and should be adjusted to indicate milliseconds. The following represents the calculation for this metric:

$$\text{SRD} = \text{Time of Status Indicative Response} - \text{Time of INVITE}$$

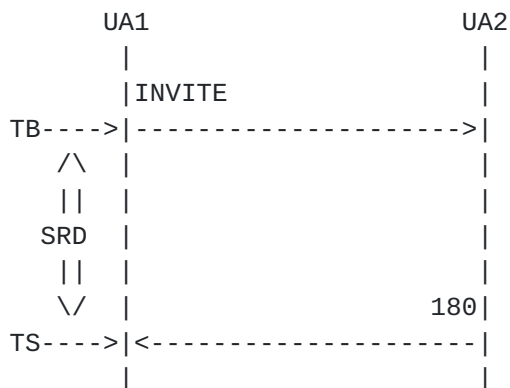
An average is one of the uses of this metric. The following represents the calculation for this metric as an average:

$$\text{ASRD} = \frac{\text{SUM (Time of Status Indicative Response - Time of INVITE)}}{\text{SUM \# of INVITE Requests}}$$

#### 3.2.1. Successful Session Setup SRD

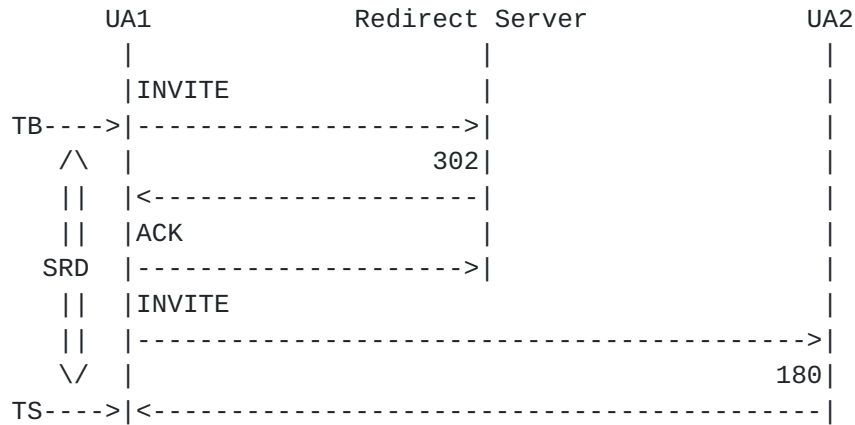
In a successful request attempt, SRD is defined as the time interval from the moment the INVITE message containing the necessary information is passed by the originating agent or user to the intended mediation or destination agent until the first provisional response is received indicating an audible or visual status of the initial session setup request. In SIP, the message indicating status would be a non-100 Trying provisional message received in response to an INVITE request. In some cases, a non-100 Trying provisional 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.

The following flow provides an example of identifiable events necessary for inputs in calculating SRD during a successful session setup request without a redirect (i.e. 3XX message):





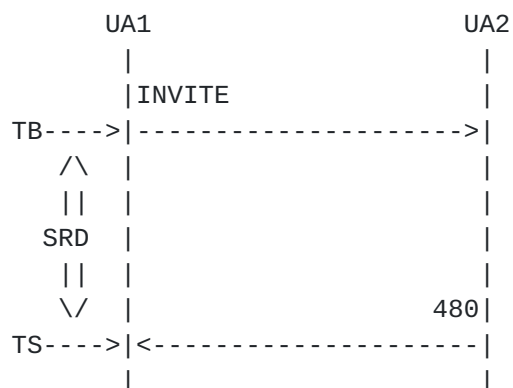
The following flow provides an example of identifiable events necessary for inputs in calculating SRD during a successful session setup with a redirect (e.g. 302 Moved Temporarily):



### 3.2.2. Failed Session Setup SRD

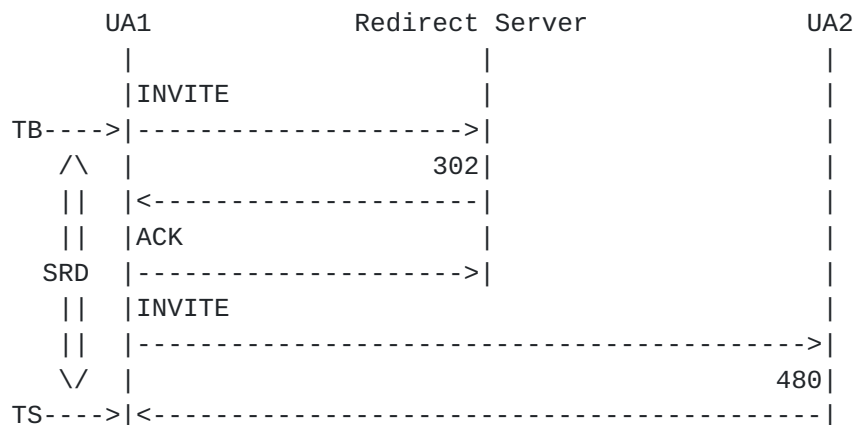
In a failed request attempt, the interval is defined from the initial session request and a non-100 Trying provisional message or a failure indication response. A failure response is described as a 4XX, 5XX, or possible 6XX message. SRD may be used to detect problems in downstream signaling functions, which may be impairing the INVITE message from reaching the intended UA.

The following flow provides an example of identifiable events necessary for inputs in calculating SRD during a failed session setup attempt without a redirect (i.e. 3XX message):



The following flow provides an example of identifiable events necessary for inputs in calculating SRD during a failed session setup attempt with a redirect (e.g. 302 Moved Temporarily):

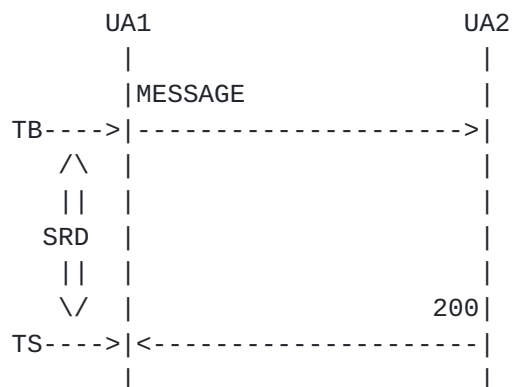




### 3.2.3. Instant Messaging

This metric is also applicable to MESSAGE [10] requests. In the above metric, INVITE can be replaced with MESSAGE to provide SRD for instant messaging (IM). The dialog will vary slightly as described in [10]. The inputs for this metric should be utilized regardless of whether a prior SIP dialog was utilized to setup the session. In that case both the SIP dialog and the MESSAGE requests are measured independently.

The following flow provides an example of identifiable events necessary for inputs in calculating SRD during a successful session MESSAGE request:



Failure requests occur similarly as to those described in [section 3.2.2](#) with MESSAGE in replacement of INVITE as the IM session request method.

### 3.3. Session Disconnect Delay (SDD)

This metric is utilized to detect failures or impairments delaying the time necessary to end a session. It can be measured from both a UAC and UAS perspective. SDD is measured for both successful and



failed session completions. The output value of this metric is numerical and should be adjusted to indicate milliseconds. The following represents the calculation for this metric:

$$\text{SDD} = \text{Time of 2XX or Timeout} - \text{Time of Completion Message (BYE)}$$

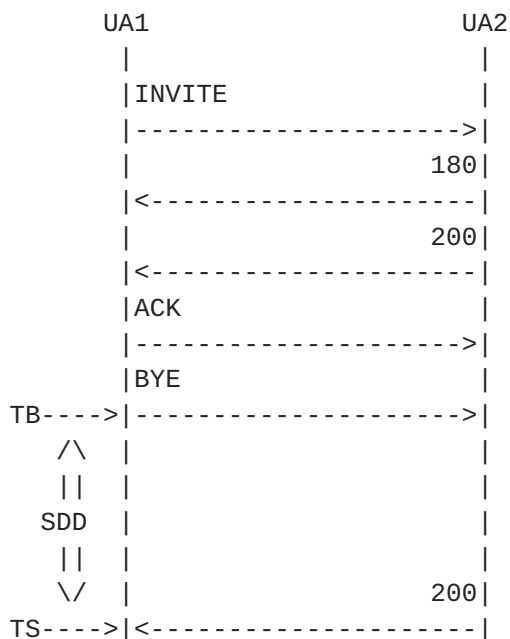
An average is one of the uses of this metric. The following represents the calculation for this metric as an average:

$$\text{ASDD} = \frac{\text{SUM (Time of 2XX or Timeout - Time of Completion Message)}}{\text{SUM \# of Completed Sessions}}$$

### **3.3.1. Successful session completion SDD**

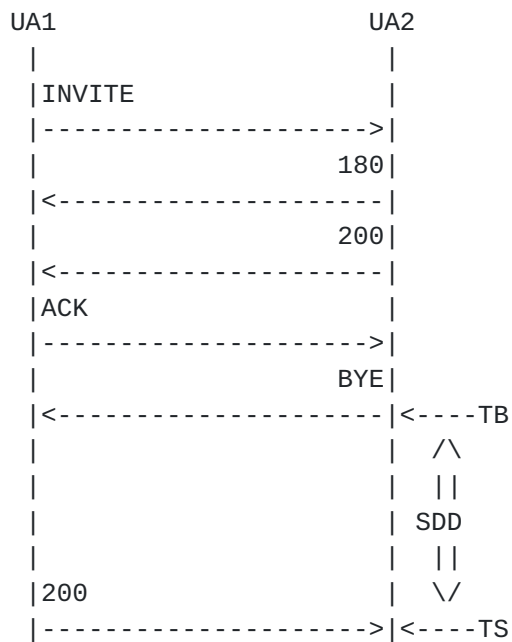
In a successful session completion, SDD is defined as the interval between sending a session completion message, such as a BYE, and receiving the subsequent 2XX acknowledgement. The following flows provide an example of identifiable events necessary for inputs in calculating SDD during a successful session completion:

Measuring SDD at the UAC -





Measuring SDD at the UAS -



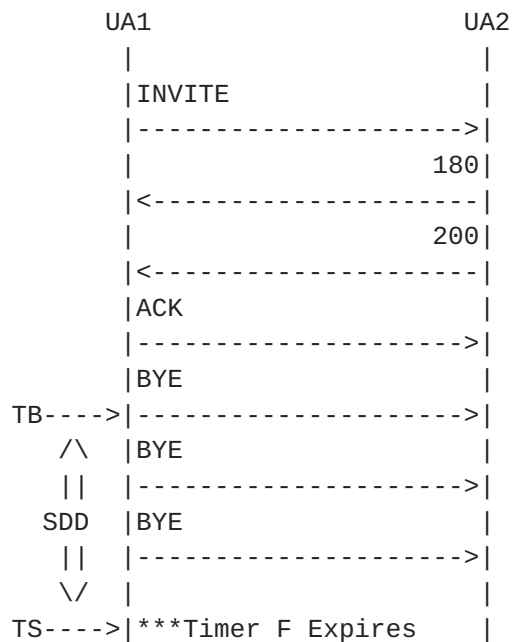
(In these two examples, TB and TS are the same even if the UAC/UAS receives the indicated messages instead of sending them.)

### **3.3.2. Failed session completion SDD**

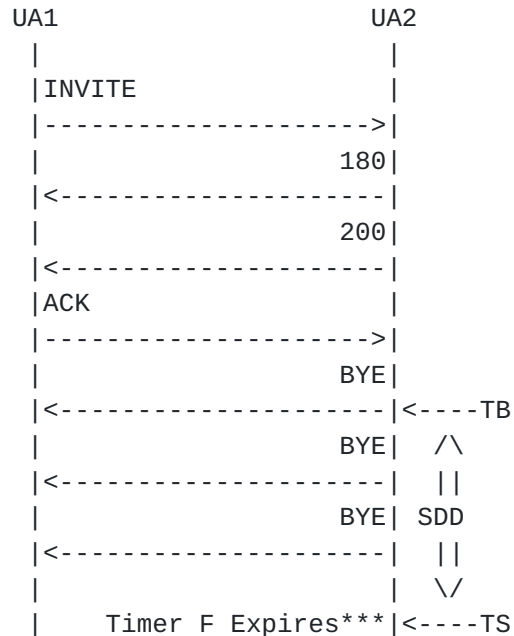
In some cases, no response is received after a session completion message is sent and potentially retried. In this case, SDD is defined as the interval between sending a session completion message, such as a BYE, and the resulting Timer F expiration. The following flows provide an example of identifiable events necessary for inputs in calculating SDD during a failed session completion attempt:



## Measuring SDD at the UAC -



## Measuring SDD at the UAS -

**3.4. Session Duration Time (SDT)**

This metric is used to detect problems (e.g. poor audio quality) causing short session durations. SDT is measured for both successful and failed session completions. It can be measured from both a UAC and UAS perspective. This metric is similar to Call Hold Time, and

is traditionally calculated as Average Call Hold Time (ACHT) in

Malas

Expires June 6, 2008

[Page 12]

telephony applications of SIP. The output value of this metric is numerical and should be adjusted to indicate minutes and seconds. The following represents the calculation for this metric:

$$\text{SDT} = \text{Time of BYE or Timeout} - \text{Time of 200 OK response to INVITE}$$

An average is one of the uses of this metric. The following represents the calculation for this metric as an average:

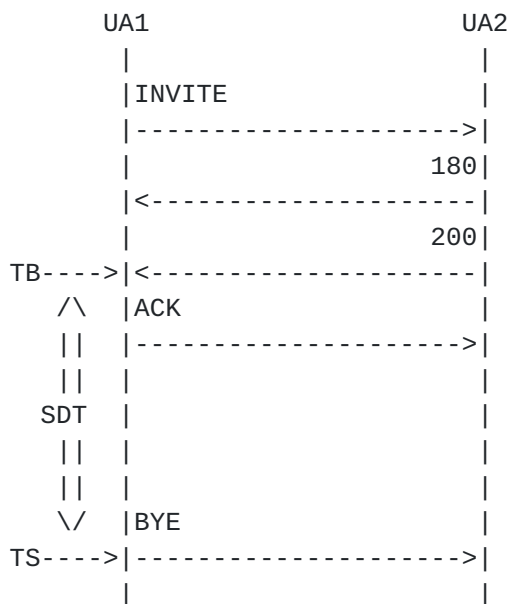
$$\text{ASDT} = \frac{\text{SUM (Time of BYE or Timeout - Time of 200 OK response to INVITE)}}{\text{SUM \# of INVITE w/ 200OK \& BYE or Timeout}}$$

#### **3.4.1. Successful session completion SDT**

In a successful session completion, SDT is calculated as an average and is defined as the duration of a dialog from receipt of a 200 OK response to an INVITE and an associated BYE message indicating dialog completion.

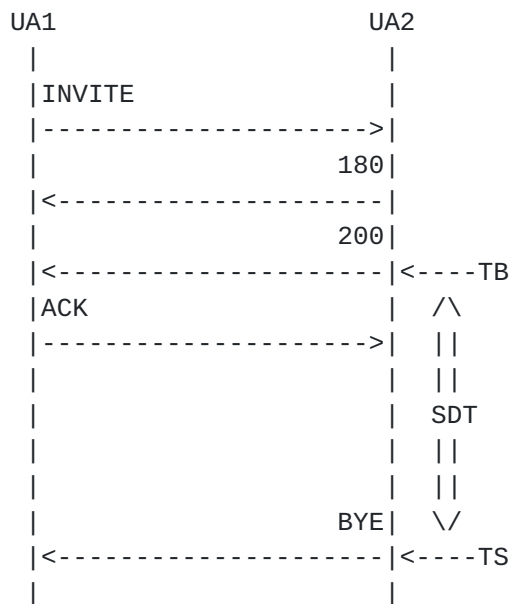
The following flows provide an example of identifiable events necessary for inputs in calculating SDT during a successful session completion (The message flows are changed between the UAC and UAS to provide varying examples.):

Measuring SDT at the UAC -





Measuring SDT at the UAS -



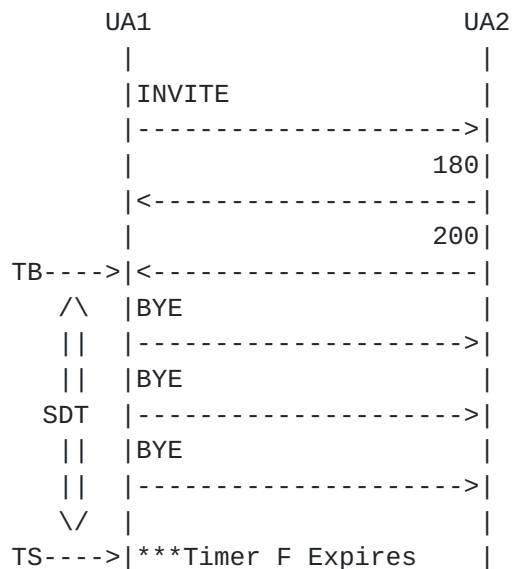
(In these two examples, TS is the same even if the UAC/UAS receives the BYE instead of sending it.)

#### **3.4.2. Failed session completion SDT**

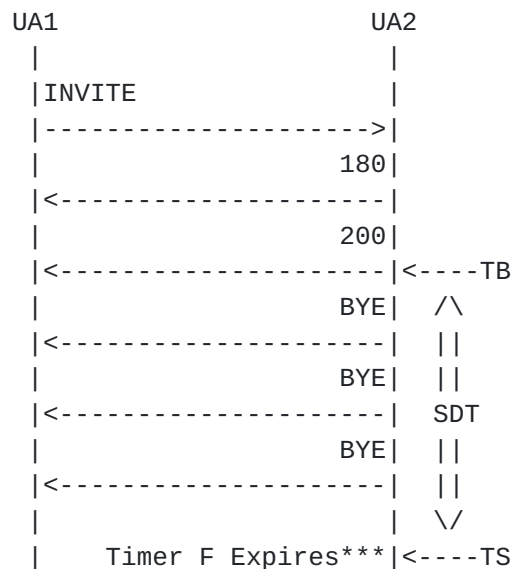
In some cases, no response is received after a session completion message is sent and potentially retried. In this case, SDT is defined as the interval between sending a session completion message, such as a BYE, and the resulting Timer F expiration. The following flows provide an example of identifiable events necessary for inputs in calculating SDT during a failed session completion attempt:



Measuring SDT at the UAC -



Measuring SDT at the UAS -



### [3.5. Average Hops per Request \(AHR\)](#)

This metric is used to indicate potential inefficient routing and to detect failure occurrences related to the number of elements traversed by a single SIP INVITE or MESSAGE request. AHR is defined as the number of hops traversed by an INVITE or MESSAGE request. It is calculated as an average. This metric requires the Max-Forwards value to be captured at both the originating UAC or proxy and the terminating UAS or proxy perspective as relative to the end-to-end



network under measurement. The output value of this metric is measured in a numerical value indicating a number of hops.

Variables =

a = Initial INVITE/MESSAGE "Max-Forwards" value

b = Initial INVITE/MESSAGE received by terminating UAS "Max-Forwards" value

c = # of Hops for INVITE/MESSAGE requests

d = SUM # of INVITE/MESSAGE requests

c = a - b

This metric is calculated as an average. The following represents the calculation for this metric:

AHR = (SUM of aggregate c's / d)

The following dialog provides an example describing the inputs necessary for this calculation. Although this example is of an INVITE SIP dialog request, a MESSAGE request is similar in its use of the Max-Forwards header. (The dialog continuation was omitted for clarity):

UA1	Proxy 1	Proxy 2	UA2
INVITE			
----->			
407			
<-----			
ACK			
----->			
INVITE (F4)			
----->	INVITE (F5)		
100		INVITE (F6)	
<-----	100	----->	
	<-----		

Message Details (Only the message details of the INVITE messages have been included for clarity. Also, some headers after Max-Forwards have been omitted for additional clarity.):



(F4) INVITE UA1 -> Proxy 1

```
INVITE sip:ua2@biloxi.example.com SIP/2.0
Via: SIP/2.0/TCP
client.atlanta.example.com:5060;branch=z9hG4bK74bf9
Max-Forwards: 70
Route: <sip:ss1.atlanta.example.com;lr>
From: UA1 <sip:ua1@atlanta.example.com>;tag=9fxced76sl
To: UA2 <sip:ua2@biloxi.example.com>
```

(F5) INVITE Proxy 1 -> Proxy 2

```
INVITE sip:ua2@biloxi.example.com SIP/2.0
Via: SIP/2.0/TCP
ss1.atlanta.example.com:5060;branch=z9hG4bK2d4790.1
Via: SIP/2.0/TCP
client.atlanta.example.com:5060;branch=z9hG4bK74bf9
;received=192.0.2.101
Max-Forwards: 69
Record-Route: <sip:ss1.atlanta.example.com;lr>
From: UA1 <sip:ua1@atlanta.example.com>;tag=9fxced76sl
To: UA2 <sip:ua2@biloxi.example.com>
```

(F6) INVITE Proxy 2 -> UA2

```
INVITE sip:ua2@client.biloxi.example.com SIP/2.0
Via: SIP/2.0/TCP ss2.biloxi.example.com:5060;branch=z9hG4bK721e4.1
Via: SIP/2.0/TCP
ss1.atlanta.example.com:5060;branch=z9hG4bK2d4790.1
;received=192.0.2.111
Via: SIP/2.0/TCP
client.atlanta.example.com:5060;branch=z9hG4bK74bf9
;received=192.0.2.101
Max-Forwards: 68
Record-Route: <sip:ss2.biloxi.example.com;lr>,
<sip:ss1.atlanta.example.com;lr>
From: UA1 <sip:ua1@atlanta.example.com>;tag=9fxced76sl
To: UA2 <sip:ua2@biloxi.example.com>
```

### **3.6. Session Establishment Rate (SER)**

This metric is used to detect the ability of a terminating UA or downstream proxy to successfully establish sessions per INVITE request. SER is defined as the number of INVITE requests resulting in a 200 OK response, to the total number of attempted INVITE requests less INVITE requests resulting in a 3XX response. This metric is similar to Answer Seizure Rate (ASR) [8] in telephony applications of SIP. It is measured at the UAC only. The output

value of this metric is numerical and should be adjusted to indicate

Malas

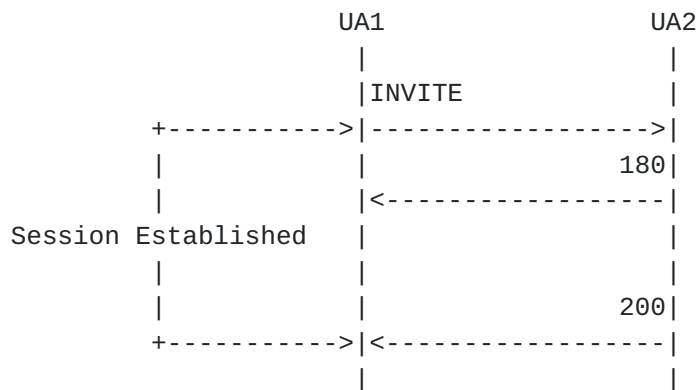
Expires June 6, 2008

[Page 17]

a percentage of successfully established sessions. The following represents the calculation for this metric:

$$\text{SER} = \frac{\text{\# of INVITE Requests w/ associated 200OK}}{(\text{Total \# of INVITE Requests}) - (\text{\# of INVITE Requests w/ 3XX Response})}$$

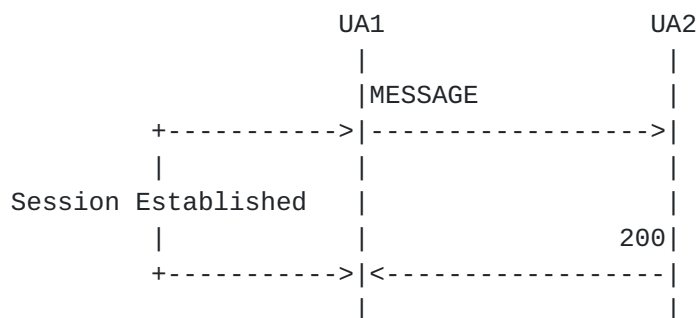
The following flow provides an example of identifiable events necessary for inputs in determining session establishment as described above:



### 3.6.1. Instant Messaging

This metric is also applicable to MESSAGE [10] requests. In the above metric, INVITE can be replaced with MESSAGE to provide SER for IM. The dialog will vary slightly as described in [10].

The following flow provides an example of identifiable events necessary for inputs in calculating SER for MESSAGE requests:



### 3.7. Session Establishment Efficiency Rate (SEER)

This metric is complimentary to SER, but is intended to exclude the potential effects of the terminating UAS from the metric. SEER is defined as the number of INVITE requests resulting in a 200 OK response and INVITE requests resulting in a 480, 486, or 600; to the total number of attempted INVITE requests less INVITE requests



resulting in a 3XX response. This metric is similar to Network Efficiency Rate (NER) [8] in telephony applications of SIP. It is measured at the UAC only. The output value of this metric is numerical and should be adjusted to indicate a percentage of successfully established sessions less common UAS failures. The following represents the calculation for this metric:

$$\text{SER} = \frac{\text{\# of INVITE Requests w/ associated 200OK, 480, 486, or 600}}{(\text{Total \# of INVITE Requests}) - (\text{\# of INVITE Requests w/ 3XX Response})}$$

Reference the example flow is [Section 3.6](#).

### **3.8. Session Defects (SD)**

Session defects provide a subset of SIP failure responses, which consistently indicate a failure in dialog processing. Defects are necessary to provide input to calculations such as Defects per Million (DPM) or other similar metrics. These failure responses are in response to initial session setup requests, such as a new INVITE. The output value of this metric is numerical and should be adjusted to indicate a percentage of defective sessions. The following failure responses provide a guideline for defective criterion:

- . 500 Server Internal Error
- . 503 Service Unavailable
- . 504 Server Timeout

This set of failure responses was derived through correlating more granular ISUP failure responses as described in [RFC 3398](#).

### **3.9. Ineffective Session Attempts (ISA)**

Ineffective session attempts occur when a proxy or agent internally releases a setup request with a failed or congested condition. This metric is similar to Ineffective Machine Attempts (IMA) in telephony applications of SIP, and was adopted from Telcordia GR-512-CORE [7]. The output value of this metric is numerical and should be adjusted to indicate a percentage of ineffective session attempts. The following failure responses provide a guideline for this criterion:

- . 408 Request Timeout
- . 500 Server Internal Error
- . 503 Service Unavailable



#### . 504 Server Timeout

This set was derived in a similar manner as described in [Section 3.6](#), in addition 408 failure responses is indicative a congested state with a downstream element.

This metric is calculated as a percentage of total session setup requests. The following represents the calculation for this metric:

$$\text{ISA \%} = \frac{\text{\# of ISA}}{\text{Total \# of Session Requests}}$$

### **[3.10. Session Disconnect Failures \(SDF\)](#)**

Session disconnect failures occur when an active session is terminated due to a failure condition that can be identified by a REASON header [5] in a BYE message. This occurs, for example, when a user agent (UA) is controlling an IP or TDM (Time Division Multiplexing) media gateway, and the media gateway notifies the UA of a failure condition causing the loss of media related to an established session. The UA will release the session with a BYE, but should include a REASON header indicating the session was disconnected abnormally. The REASON value is utilized to determine the disconnect was a failure. This metric is similar to Cutoff Calls (CC) in telephony applications of SIP, and was adopted from Telcordia GR-512-CORE [7]. The input variables for this metric are captured from the originating UAC or proxy perspective as relative to the end-to-end network under measurement. The output value of this metric is numerical and should be adjusted to indicate a percentage of session disconnect failures.

This metric is calculated as a percentage of total session completed successfully as defined in [Section 3.5](#). The following represents the calculation for this metric:

$$\text{SDF \%} = \frac{\text{\# of SDF's}}{\text{Total \# of Session Requests}}$$

### **[3.11. Session Completion Rate \(SCR\)](#)**

A session completion is defined as a SIP dialog, which completes without failing due to a lack of response from an intended proxy or UA. This metric is only used when at least one proxy is involved in the dialog. This metric is similar to Call Completion Rate (CCR) in telephony applications of SIP. The output value of this metric is numerical and should be adjusted to indicate a percentage of successfully completed sessions.



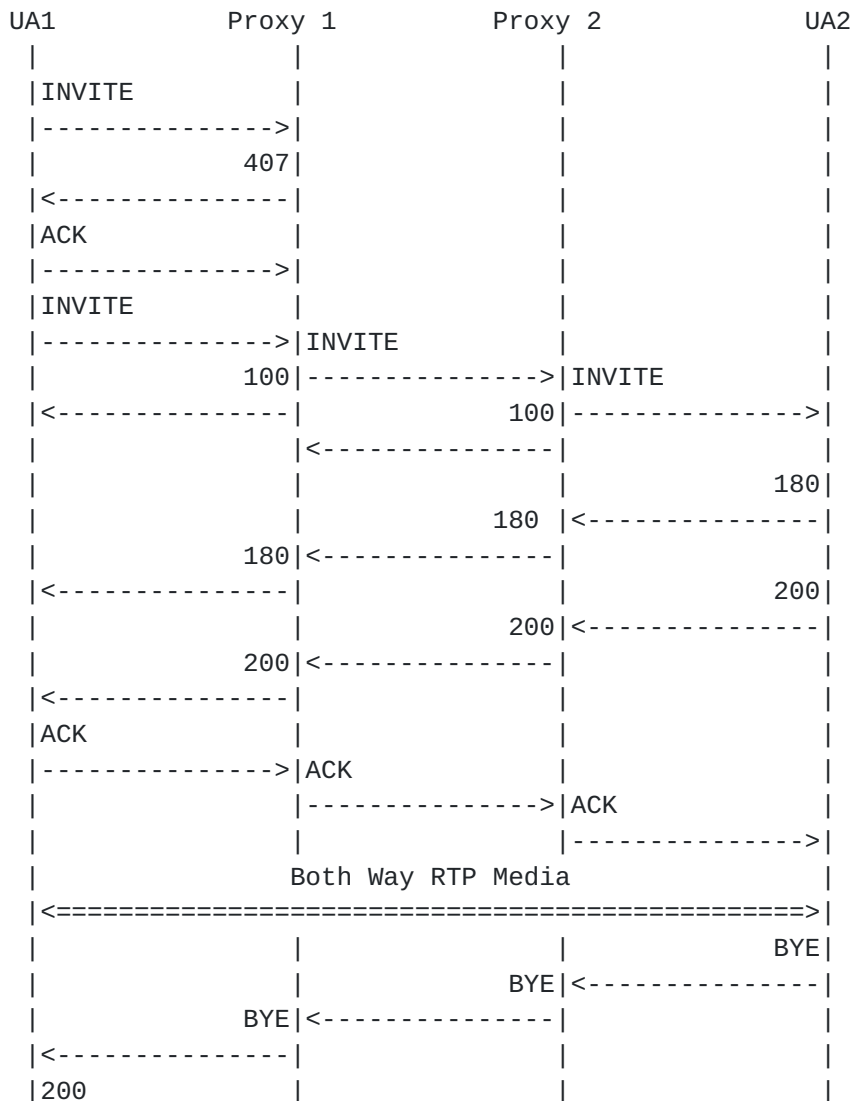
This metric is calculated as a percentage of total sessions completed successfully. The following represents the calculation for this metric:

$$\text{SCR \%} = \frac{\text{\# of Successfully Completed Sessions}}{\text{Total \# of Session Requests}}$$

### **3.11.1. Successful Session Completion**

A session completes successfully when it begins with a setup request and ends with a session completion message.

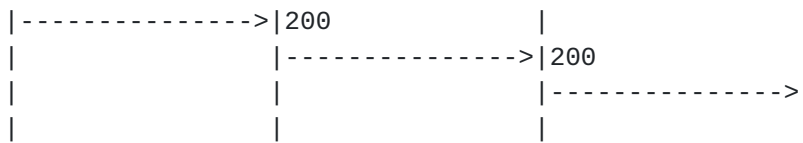
The following dialog [4] provides an example describing the necessary events of a successful session completion:



Malas

Expires June 6, 2008

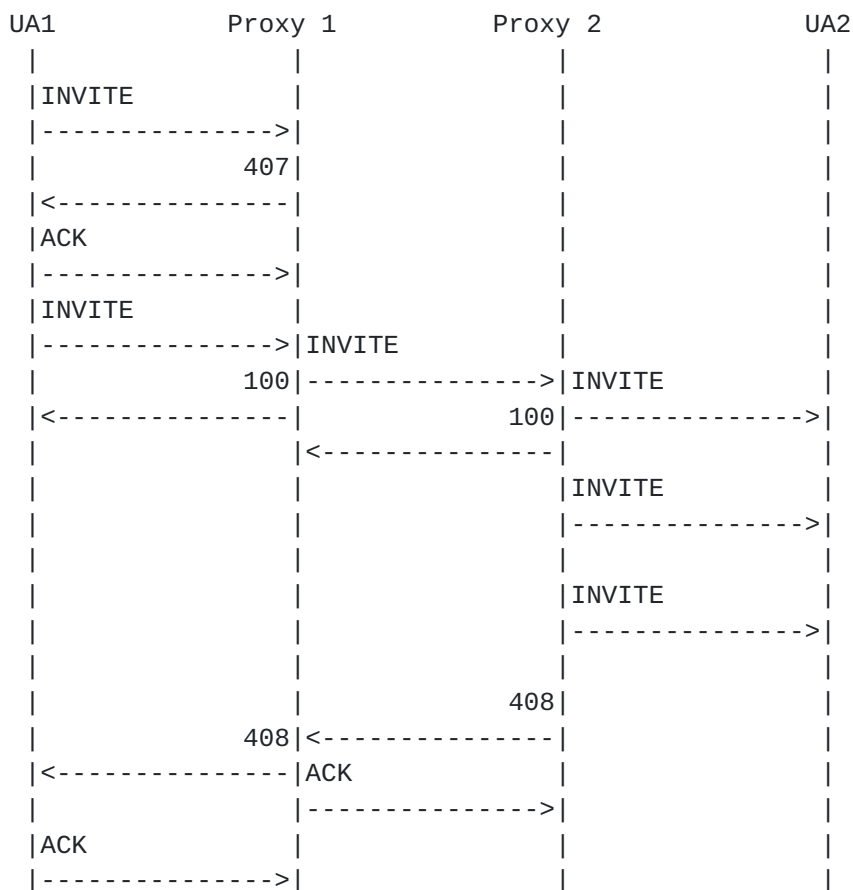
[Page 21]



### 3.11.2. Failed Session Completion

Session completion fails when an INVITE is sent from a UAC, but there is no indication the INVITE reached the intended UAS. This can also occur if the intended UAS does not respond to the UAC or the response never reaches the UAC associated with the session.

The following dialog provides an example describing the necessary events of an unsuccessful session completion:



### 3.12. Session Success Rate (SSR)

Session success rate is defined as the percentage of successfully completed sessions compared to sessions, which fail due to ISA or SDF. This metric is also known as Call Success Rate (CSR) in



telephony applications of SIP. The output value of this metric is numerical and should be adjusted to indicate a percentage of successful sessions. The following represents the calculation for this metric:

$$\text{SSR} = 100\% - (\text{ISA}\% + \text{SDF}\%)$$

#### **4. Metric Correlations**

These metrics may be used to determine the performance of a domain and/or user. This would be to provide a metric relative to one or more dimensions. The following is a subset of dimensions for providing further granularity per metric:

- . To "user"
- . From "user"
- . Bi-direction "user"
- . To "domain"
- . From "domain"
- . Bi-direction "domain"

#### **5. Additional Considerations**

##### **5.1. Back-to-back User Agent (B2BUA)**

A B2BUA may impact the ability to collect these metrics with an end-to-end perspective. It is necessary to realize a B2BUA may act as an originating UAC and terminating UAS or it may act as a proxy. In some cases, it may be necessary to consider information collected from both sides of the B2BUA in order to determine the end-to-end perspective. In other cases, the B2BUA may act simply as a proxy allowing data to be derived as necessary for the input into any of the listed calculations.

##### **5.2. Authorization and Authentication**

During the process of setting up a SIP dialog, various authentication methods may be utilized. These authentication methods will add to the duration as measured by the metrics, and the length of time will vary based on those methods. The failures of these authentication methods will also be captured by these metrics, since SIP is ultimately used to indicate the success or failure of the authorization and/or authentication attempt. The metrics in [section 3](#) are inclusive of the duration associated with this process, even if



the method is external to the SIP protocol. This was included purposefully, due to its inherent impact on the protocol and the subsequent SIP dialogs.

### **5.3. Forking**

Forking should be considered when determining the messages associated with the input values for the described metrics. If all of the forked dialogs were used in the metric calculations, the numbers would skew dramatically. There are two different points of forking, which must be considered. First, forking may occur at a proxy downstream from the UAC that is being used for metric input values. Since, the downstream proxy is responsible for forking a message and then only sending the accepted response to the UAC, the UAC will only see messages as indicated in the described metrics. Second, in the cases where the observed UAC or proxy is forking the messages, then it must utilize the first INVITE or set of INVITE messages sent and the first accepted 200 OK. A tag will identify this dialog as distinct from the other 200 OK responses, which are acknowledged and an immediate BYE is sent. The application responsible for capturing and/or understanding the input values must utilize this tag to distinguish between dialogs.

### **5.4. Data Collection**

The input necessary for these calculations may be collected in a number of different manners. It may be collected or retrieved from call detail records (CDR) or raw signaling information generated by a proxy or UA. When using records, time synchronization must be considered between applicable elements.

The information may also be transmitted through the use of network management protocols like Simple Network Management Protocol (SNMP) [[RFC 1157](#)] and via future extensions to the SIP Management Information Base (MIB) modules [[6](#)], or through a potential undefined new performance metric event package [[3](#)] retrieved via SUBSCRIBE requests.

Data may be collected for a sample of calls or all calls, and may also be derived from test call scenarios. These metrics are flexible based on the needs of the application.

### **5.5. Testing Documentation**

In some cases, these metrics will be used to provide output values to signify the performance level of a specific SIP-based element. When using these metrics in a test environment, the environment must be accurately documented for the purposes of replicating any output values in future testing and/or validation.



### **5.6. Metric Template**

Although, this document provides details for a foundational set of pertinent metrics, other metrics may be necessary as the SIP protocol evolves or as deemed necessary by an individual or set of companies and/or vendors. This section details a template, which should be used when creating new performance metrics. This template will allow for a common structure of information, which will enable a more adaptable method of understanding and incorporating metrics defined beyond the scope of this document.

All metrics should include the following characteristics:

- . A metric should have a common 2-4 word summary description, which can be identified as a 2-4 letter acronym.
- . The metric must describe the problem or motivation for which it is attempting to detect or identify.
- . The metric must describe what is measured as indicated specifically by defined SIP messages.
- . The metric must identify the unit(s) of measure described in the associated output.
- . The metric must define the time at which the inputs are captured, including both beginning and end.
- . The metric must describe if the outputs can be utilized in a manner other than the raw output (e.g. average, high/low, etc.), and if so, how.

## **6. 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].

## **7. IANA Considerations**

There are no IANA considerations at this time.

## **8. 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; however, they may also be used for testing end-to-end SIP-based service environments.

## **9. Contributors**

The following people made substantial contributions to this work:

Carol Davids	Illinois Institute of Technology
Al Morton	AT&T Labs
Marian Delkinov	Ericsson
Adam Uzelac	Global Crossing
Jean-Francois Mule	CableLabs
Rich Terpstra	Level 3 Communications

## **10. Acknowledgments**

I would like to thank John Hearty and Dean Bayless for their efforts in reviewing the draft and providing insight regarding clarification of certain aspects described throughout the draft.

## **11. References**

### **11.1. Normative References**

- [1] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [2] 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.
- [3] Roach, A., "Session Initiation Protocol (SIP)-Specific Event Notification", [RFC 3265](#), June 2002.
- [4] Johnston, A., Donovan, S., Sparks, R., Cunningham, C., and K. Summers, "Session Initiation Protocol (SIP) Basic Call Flow Examples", [BCP 75](#), [RFC 3665](#), December 2003.
- [5] Schulzrinne, H., Oran, D., Camarillo, G., "The Reason Header Field for the Sessions Initiation Protocol (SIP)", [RFC 3326](#), December 2002.
- [6] Lingle, K., Mule, J., Maeng, J., Walker, D., "Management Information Base for the Session Initiation Protocol (SIP)", [draft-ietf-sip-mib-10](#), Work in Progress.
- [7] Telcordia, "LSSGR: Reliability, [Section 12](#)", GR-512-CORE, Issue 2, January 1998.



- [8] ITU-T, "Series E: Overall Network Operation, Telephone Service, Service Operation and Human Factors", E.411, March 2000.
- [9] ITU-T, "Series E: Overall Network Operation, Telephone Service, Service Operation and Human Factors", E.721, May 1999.
- [10] B. Campbell, Ed, Rosenberg, J., Schulzrinne, H., Huitema, C., Gurle, D., "Session Initiation Protocol (SIP) Extension for Instant Messaging", [RFC 3428](#), December 2002.

## **11.2. 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](mailto: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, THE IETF TRUST 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 IETF Trust (2007).

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.

