

Benchmarking Methodology Working  
Group  
Internet-Draft  
Expires: May 12, 2013

C. Davids  
Illinois Institute of Technology  
V. Gurbani  
Bell Laboratories, Alcatel-Lucent  
S. Poretsky  
Allot Communications  
November 8, 2012

**Methodology for Benchmarking SIP Networking Devices**  
**draft-ietf-bmwg-sip-bench-meth-06**

**Abstract**

This document describes the methodology for benchmarking Session Initiation Protocol (SIP) performance as described in SIP benchmarking terminology document. The methodology and terminology are to be used for benchmarking signaling plane performance with varying signaling and media load. Both scale and establishment rate are measured by signaling plane performance. The SIP Devices to be benchmarked may be a single device under test (DUT) or a system under test (SUT). Benchmarks can be obtained and compared for different types of devices such as SIP Proxy Server, SBC, and server paired with a media relay or Firewall/NAT device.

**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 May 12, 2013.

**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

<a href="#">1.</a>	<a href="#">Terminology</a>	<a href="#">4</a>
<a href="#">2.</a>	<a href="#">Introduction</a>	<a href="#">4</a>
<a href="#">3.</a>	<a href="#">Benchmarking Topologies</a>	<a href="#">5</a>
<a href="#">4.</a>	<a href="#">Test Setup Parameters</a>	<a href="#">5</a>
<a href="#">4.1.</a>	<a href="#">Selection of SIP Transport Protocol</a>	<a href="#">5</a>
<a href="#">4.2.</a>	<a href="#">Signaling Server</a>	<a href="#">5</a>
<a href="#">4.3.</a>	<a href="#">Associated Media</a>	<a href="#">5</a>
<a href="#">4.4.</a>	<a href="#">Selection of Associated Media Protocol</a>	<a href="#">6</a>
<a href="#">4.5.</a>	<a href="#">Number of Associated Media Streams per SIP Session</a>	<a href="#">6</a>
<a href="#">4.6.</a>	<a href="#">Session Duration</a>	<a href="#">6</a>
<a href="#">4.7.</a>	<a href="#">Attempted Sessions per Second</a>	<a href="#">6</a>
<a href="#">4.8.</a>	<a href="#">Stress Testing</a>	<a href="#">6</a>
<a href="#">4.9.</a>	<a href="#">Benchmarking algorithm</a>	<a href="#">6</a>
<a href="#">5.</a>	<a href="#">Reporting Format</a>	<a href="#">9</a>
<a href="#">5.1.</a>	<a href="#">Test Setup Report</a>	<a href="#">9</a>
<a href="#">5.2.</a>	<a href="#">Device Benchmarks for IS</a>	<a href="#">10</a>
<a href="#">5.3.</a>	<a href="#">Device Benchmarks for NS</a>	<a href="#">10</a>
<a href="#">6.</a>	<a href="#">Test Cases</a>	<a href="#">10</a>
<a href="#">6.1.</a>	<a href="#">Baseline Session Establishment Rate of the test bed</a>	<a href="#">10</a>
<a href="#">6.2.</a>	<a href="#">Session Establishment Rate without media</a>	<a href="#">11</a>
<a href="#">6.3.</a>	<a href="#">Session Establishment Rate with Media on DUT/SUT</a>	<a href="#">11</a>
<a href="#">6.4.</a>	<a href="#">Session Establishment Rate with Media not on DUT/SUT</a>	<a href="#">12</a>
<a href="#">6.5.</a>	<a href="#">Session Establishment Rate with Loop Detection Enabled</a>	<a href="#">13</a>
<a href="#">6.6.</a>	<a href="#">Session Establishment Rate with Forking</a>	<a href="#">13</a>
<a href="#">6.7.</a>	<a href="#">Session Establishment Rate with Forking and Loop Detection</a>	<a href="#">14</a>
<a href="#">6.8.</a>	<a href="#">Session Establishment Rate with TLS Encrypted SIP</a>	<a href="#">14</a>
<a href="#">6.9.</a>	<a href="#">Session Establishment Rate with IPsec Encrypted SIP</a>	<a href="#">15</a>
<a href="#">6.10.</a>	<a href="#">Session Establishment Rate with SIP Flooding</a>	<a href="#">15</a>
<a href="#">6.11.</a>	<a href="#">Maximum Registration Rate</a>	<a href="#">16</a>
<a href="#">6.12.</a>	<a href="#">Maximum Re-Registration Rate</a>	<a href="#">16</a>
<a href="#">6.13.</a>	<a href="#">Maximum IM Rate</a>	<a href="#">17</a>
<a href="#">6.14.</a>	<a href="#">Session Capacity without Media</a>	<a href="#">17</a>
<a href="#">6.15.</a>	<a href="#">Session Capacity with Media</a>	<a href="#">18</a>
<a href="#">6.16.</a>	<a href="#">Session Capacity with Media and a Media Relay/NAT and/or Firewall</a>	<a href="#">18</a>
<a href="#">7.</a>	<a href="#">IANA Considerations</a>	<a href="#">19</a>
<a href="#">8.</a>	<a href="#">Security Considerations</a>	<a href="#">19</a>
<a href="#">9.</a>	<a href="#">Acknowledgments</a>	<a href="#">19</a>
<a href="#">10.</a>	<a href="#">References</a>	<a href="#">19</a>
<a href="#">10.1.</a>	<a href="#">Normative References</a>	<a href="#">19</a>
<a href="#">10.2.</a>	<a href="#">Informative References</a>	<a href="#">20</a>
	<a href="#">Authors' Addresses</a>	<a href="#">20</a>



## 1. Terminology

In this document, the key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" are to be interpreted as described in [BCP 14](#), conforming to [\[RFC2119\]](#) and indicate requirement levels for compliant implementations.

Terms specific to SIP [\[RFC3261\]](#) performance benchmarking are defined in [\[I-D.sip-bench-term\]](#).

[RFC 2119](#) defines the use of these key words to help make the intent of standards track documents as clear as possible. While this document uses these keywords, this document is not a standards track document. The term Throughput is defined in [\[RFC2544\]](#).

## 2. Introduction

This document describes the methodology for benchmarking Session Initiation Protocol (SIP) performance as described in Terminology document [\[I-D.sip-bench-term\]](#). The methodology and terminology are to be used for benchmarking signaling plane performance with varying signaling and media load. Both scale and establishment rate are measured by signaling plane performance.

The SIP Devices to be benchmarked may be a single device under test (DUT) or a system under test (SUT). The DUT is a SIP Server, which may be any [\[RFC3261\]](#) conforming device. The SUT can be any device or group of devices containing [RFC 3261](#) conforming functionality along with Firewall and/or NAT functionality. This enables benchmarks to be obtained and compared for different types of devices such as SIP Proxy Server, SBC, SIP proxy server paired with a media relay or Firewall/NAT device. SIP Associated Media benchmarks can also be made when testing SUTs.

The test cases covered in this methodology document provide benchmarks metrics of Registration Rate, SIP Session Establishment Rate, Session Capacity, and IM Rate. These can be benchmarked with or without associated Media. Some cases are also included to cover Forking, Loop detection, Encrypted SIP, and SIP Flooding. The test topologies that can be used are described in the Test Setup section. Topologies are provided for benchmarking of a DUT or SUT. Benchmarking with Associated Media can be performed when using a SUT.

SIP permits a wide range of configuration options that are also explained in the Test Setup section. Benchmark metrics could possibly be impacted by Associated Media. The selected values for



Session Duration and Media Streams Per Session enable benchmark metrics to be benchmarked without Associated Media. Session Setup Rate could possibly be impacted by the selected value for Maximum Sessions Attempted. The benchmark for Session Establishment Rate is measured with a fixed value for maximum Session Attempts.

Finally, the overall value of these tests is to serve as a comparison function between multiple SIP implementations. One way to use these tests is to derive benchmarks with SIP devices from Vendor-A, derive a new set of benchmarks with similar SIP devices from Vendor-B and perform a comparison on the results of Vendor-A and Vendor-B. This document does not make any claims on the interpretation of such results.

### **3. Benchmarking Topologies**

Familiarity with the benchmarking models in Section 2.2 of [\[I-D.sip-bench-term\]](#) is assumed. Figures 1 through 10 in [\[I-D.sip-bench-term\]](#) contain the canonical topologies that can be used to perform the benchmarking tests listed in this document.

### **4. Test Setup Parameters**

#### **4.1. Selection of SIP Transport Protocol**

Test cases may be performed with any transport protocol supported by SIP. This includes, but is not limited to, SIP TCP, SIP UDP, and TLS. The protocol used for the SIP transport protocol must be reported with benchmarking results.

#### **4.2. Signaling Server**

The Signaling Server is defined in the companion terminology document, ([\[I-D.sip-bench-term\]](#), Section 3.2.2) It is a SIP-speaking device that complies with [RFC 3261](#). Conformance to [\[RFC3261\]](#) is assumed for all tests. The Signaling Server may be the DUT or a component of a SUT. The Signaling Server may include Firewall and/or NAT functionality. The components of the SUT may be a single physical device or separate devices.

#### **4.3. Associated Media**

Some tests require Associated Media to be present for each SIP session. The test topologies to be used when benchmarking SUT performance for Associated Media are shown in [\[I-D.sip-bench-term\]](#), Figures 4 and 5.





#### **4.4. Selection of Associated Media Protocol**

The test cases specified in this document provide SIP performance independent of the protocol used for the media stream. Any media protocol supported by SIP may be used. This includes, but is not limited to, RTP, RTSP, and SRTP. The protocol used for Associated Media MUST be reported with benchmarking results.

#### **4.5. Number of Associated Media Streams per SIP Session**

Benchmarking results may vary with the number of media streams per SIP session. When benchmarking a SUT for voice, a single media stream is used. When benchmarking a SUT for voice and video, two media streams are used. The number of Associated Media Streams MUST be reported with benchmarking results.

#### **4.6. Session Duration**

SUT performance benchmarks may vary with the duration of SIP sessions. Session Duration MUST be reported with benchmarking results. A Session Duration of zero seconds indicates transmission of a BYE immediately following successful SIP establishment indicate by receipt of a 200 OK. An infinite Session Duration indicates that a BYE is never transmitted.

#### **4.7. Attempted Sessions per Second**

DUT and SUT performance benchmarks may vary with the the rate of attempted sessions offered by the Tester. Attempted Sessions per Second MUST be reported with benchmarking results.

#### **4.8. Stress Testing**

The purpose of this document is to benchmark SIP performance; this document does not benchmark stability of SIP systems under stressful conditions such as a high rate of Attempted Sessions per Second.

#### **4.9. Benchmarking algorithm**

In order to benchmark the test cases uniformly in [Section 6](#), the algorithm described in this section should be used. Both, a prosaic description of the algorithm and a pseudo-code description are provided.

The goal is to find the largest value of a SIP session-request-rate, measured in sessions-per-second, which the DUT/SUT can process with zero errors. To discover that number, an iterative process (defined below) is used to find a candidate for this rate. Once the candidate



rate has been found, the DUT/SUT is subjected to an offered load whose arrival rate is set to that of the candidate rate. This test is run for an extended period of time, which is referred to as infinity, and which is, itself, a parameter of the test labeled T in the pseudo-code. This latter phase of testing is called the steady-state phase. If errors are encountered during this steady-state phase, then the candidate rate is reduced by a defined percent, also a parameter of test, and the steady-state phase is entered again until a final (new) steady-state rate is achieved.

The iterative process itself is defined as follows: a starting rate of 100 sessions per second (sps) is selected. The test is executed for the time period identified by t in the pseudo-code below. If no failures occur, the rate is increased to 150 sps and again tested for time period t. The attempt rate is continuously ramped up until a failure is encountered before the end of the test time t. Then an attempt rate is calculated that is higher than the last successful attempt rate by a quantity equal to half the difference between the rate at which failures occurred and the last successful rate. If this new attempt rate also results in errors, a new attempt rate is tried that is higher than the last successful attempt rate by a quantity equal to half the difference between the rate at which failures occurred and the last successful rate. Continuing in this way, an attempt rate without errors is found. The operator can specify margin of error using the parameter G, measured in units of sessions per second.

The pseudo-code corresponding to the description above follows.

```
; ---- Parameters of test, adjust as needed
t  := 5000    ; local maximum; used to figure out largest
              ; value
T  := 50000   ; global maximum; once largest value has been
              ; figured out, pump this many requests before calling
              ; the test a success
m  := {...}   ; other attributes that affect testing, such
              ; as media streams, etc.
s  := 100     ; Initial session attempt rate (in sessions/sec)
G  := 5       ; granularity of results - the margin of error in sps
C  := 0.05    ; calibration amount: How much to back down if we
              ; have found candidate s but cannot send at rate s for
              ; time T without failures

; ---- End of parameters of test
; ---- Initialization of flags, candidate values and upper bounds

f  := false   ; indicates that you had a success after the upper limit
```



```
F := false ; indicates that test is done
c := 0      ; indicates that we have found an upper limit

proc main
  find_largest_value ; First, figure out the largest value.

  ; Now that the largest value (saved in s) has been figured out,
  ; use it for sending out s requests/s and send out T requests.

  do {
    send_traffic(s, m, T) ; send_traffic not shown
    if (all requests succeeded) {
      F := true ; test is done
    } else if (one or more requests fail) {
      s := s - (C * s) ; Reduce s by calibration amount
      steady_state
    }
  } while (F == false)
end proc

proc find_largest_value
  ; Iterative process to figure out the largest value we can
  ; handle with no failures
  do {
    send_traffic(s, m, t) ; Send s request/sec with m
                          ; characteristics until t requests have
                          ; been sent
    if (all requests succeeded) {
      s' := s ; save candidate value of metric

      if ( c == 0 ) {
        s := s + (0.5 * s)

      }else if ((c == 1) && (s''-s')) > 2*G ) {
        s := s + ( 0.5 * (s'' - s ));

      }else if ((c == 1) && ((s''-s') <= 2*G ) {
        f := true;

      }

      else if (one or more requests fail) {
        c := 1 ; we have found an upper bound for the metric
        s'' := s ; save new upper bound
        s := s - (0.5 * (s - s'))
      }
    } while (f == false)
  end proc
```



## 5. Reporting Format

### 5.1. Test Setup Report

SIP Transport Protocol = \_\_\_\_\_  
(valid values: TCP|UDP|TLS|SCTP|specify-other)  
Session Attempt Rate = \_\_\_\_\_  
(session attempts/sec)  
IS Media Attempt Rate = \_\_\_\_\_  
(IS media attempts/sec)  
Total Sessions Attempted = \_\_\_\_\_  
(total sessions to be created over duration of test)  
Media Streams Per Session = \_\_\_\_\_  
(number of streams per session)  
Associated Media Protocol = \_\_\_\_\_  
(RTP|RTSP|specify-other)  
Media Packet Size = \_\_\_\_\_  
(bytes)  
Media Offered Load = \_\_\_\_\_  
(packets per second)  
Media Session Hold Time = \_\_\_\_\_  
(seconds)  
Establishment Threshold time = \_\_\_\_\_  
(seconds)  
Loop Detecting Option = \_\_\_\_\_  
(on|off)  
Forking Option  
    Number of endpoints request sent to = \_\_\_\_\_  
    (1, means forking is not enabled)  
    Type of forking = \_\_\_\_\_  
    (serial|parallel)  
Authentication option = \_\_\_\_\_  
    (on|off; if on, please see Notes 2 and 3 below).

Note 1: Total Sessions Attempted is used in the calculation of the Session Establishment Performance ([[I-D.sip-bench-term](#)], [Section 3.4.5](#)). It is the number of session attempts ([[I-D.sip-bench-term](#)], [Section 3.1.6](#)) that will be made over the duration of the test.

Note 2: When the Authentication Option is "on" the test tool must be set to ignore 401 and 407 failure responses in any test described as a "test to failure." If this is not done, all such tests will yield trivial benchmarks, as all attempt rates will lead to a failure after the first attempt.

Note 3: When the Authentication Option is "on" the DUT/SUT uses two





transactions instead of one when it is establishing a session or accomplishing a registration. The first transaction ends with the 401 or 407. The second ends with the 200 OK or another failure message. The Test Organization interested in knowing how many times the EA was intended to send a REGISTER as distinct from how many times the EA wound up actually sending a REGISTER may wish to record the following data as well: Number of responses of the following type: 401: \_\_\_\_\_ (if authentication turned on; N/A otherwise)  
407: \_\_\_\_\_ (if authentication turned on; N/A otherwise)

## **5.2. Device Benchmarks for IS**

Registration Rate = \_\_\_\_\_  
(registrations per second)  
Re-registration Rate = \_\_\_\_\_  
(registrations per second)  
Session Capacity = \_\_\_\_\_  
(sessions)  
Session Overload Capacity = \_\_\_\_\_  
(sessions)  
Session Establishment Rate = \_\_\_\_\_  
(sessions per second)  
Session Establishment Performance = \_\_\_\_\_  
(total established sessions/total sessions attempted)(no units)  
Session Attempt Delay = \_\_\_\_\_  
(seconds)

## **5.3. Device Benchmarks for NS**

IM Rate = \_\_\_\_\_ (IM messages per second)

## **6. Test Cases**

### **6.1. Baseline Session Establishment Rate of the test bed**

#### **Objective:**

To benchmark the Session Establishment Rate of the Emulated Agent (EA) with zero failures.

#### **Procedure:**

1. Configure the DUT in the test topology shown in Figure 1 in [[I-D.sip-bench-term](#)].



2. Set media streams per session to 0.
3. Execute benchmarking algorithm as defined in [Section 4.9](#) to get the baseline session establishment rate. This rate MUST be recorded using any pertinent parameters as shown in the reporting format of [Section 5.1](#).

Expected Results: This is the scenario to obtain the maximum Session Establishment Rate of the EA and the test bed when no DUT/SUT is present. The results of this test might be used to normalize test results performed on different test beds or simply to better understand the impact of the DUT/SUT on the test bed in question.

### **6.2. Session Establishment Rate without media**

#### **Objective:**

To benchmark the Session Establishment Rate of the DUT/SUT with no associated media and zero failures.

#### **Procedure:**

1. If the DUT/SUT is being benchmarked as a user agent client or a user agent server, configure the DUT in the test topology shown in Figure 1 or Figure 2 in [[I-D.sip-bench-term](#)]. Alternatively, if the DUT is being benchmarked as a proxy or a B2BUA, configure the DUT in the test topology shown in Figure 5 in [[I-D.sip-bench-term](#)].
2. Configure a SUT according to the test topology shown in Figure 7 in [[I-D.sip-bench-term](#)].
3. Set media streams per session to 0.
4. Execute benchmarking algorithm as defined in [Section 4.9](#) to get the session establishment rate. This rate MUST be recorded using any pertinent parameters as shown in the reporting format of [Section 5.1](#).

Expected Results: This is the scenario to obtain the maximum Session Establishment Rate of the DUT/SUT.

### **6.3. Session Establishment Rate with Media on DUT/SUT**

#### **Objective:**

To benchmark the Session Establishment Rate of the DUT/SUT with zero failures when Associated Media is included in the benchmark test and the media is running through the DUT/SUT.

#### **Procedure:**

1. If the DUT is being benchmarked as a user agent client or a user agent server, configure the DUT in the test topology shown in Figure 3 or Figure 4 of [[I-D.sip-bench-term](#)]. Alternatively, if the DUT is being benchmarked as a B2BUA,



- configure the DUT in the test topology shown in Figure 6 in [\[I-D.sip-bench-term\]](#).
2. Configure a SUT according to the test topology shown in Figure 9 in [\[I-D.sip-bench-term\]](#).
  3. Set media streams per session to 1.
  4. Execute benchmarking algorithm as defined in [Section 4.9](#) to get the session establishment rate with media. This rate MUST be recorded using any pertinent parameters as shown in the reporting format of [Section 5.1](#).

Expected Results: Session Establishment Rate results obtained with Associated Media with any number of media streams per SIP session are expected to be identical to the Session Establishment Rate results obtained without media in the case where the server is running on a platform separate from the platform on which the Media Relay, NAT or Firewall is running. Session Establishment Rate results obtained with Associated Media may be lower than those obtained without media in the case where the server and the NAT, Firewall or Media Relay are running on the same platform.

#### **[6.4.](#) Session Establishment Rate with Media not on DUT/SUT**

Objective:

To benchmark the Session Establishment Rate of the DUT/SUT with zero failures when Associated Media is included in the benchmark test but the media is not running through the DUT/SUT.

Procedure:

1. If the DUT is being benchmarked as proxy or B2BUA, configure the DUT in the test topology shown in Figure 7 in [\[I-D.sip-bench-term\]](#).
2. Configure a SUT according to the test topology shown in Figure 8 in [\[I-D.sip-bench-term\]](#).
3. Set media streams per session to 1.
4. Execute benchmarking algorithm as defined in [Section 4.9](#) to get the session establishment rate with media. This rate MUST be recorded using any pertinent parameters as shown in the reporting format of [Section 5.1](#).

Expected Results: Session Establishment Rate results obtained with Associated Media with any number of media streams per SIP session are expected to be identical to the Session Establishment Rate results obtained without media in the case where the server is running on a platform separate from the platform on which the Media Relay, NAT or Firewall is running. Session Establishment Rate results obtained with Associated Media may be lower than those obtained without media in the case where the server and the NAT, Firewall or Media Relay are running on the same platform.



### **6.5. Session Establishment Rate with Loop Detection Enabled**

**Objective:**

To benchmark the Session Establishment Rate of the DUT/SUT with zero failures when the Loop Detection option is enabled and no media streams are present.

**Procedure:**

1. If the DUT is being benchmarked as a proxy or B2BUA, and loop detection is supported in the DUT, then configure the DUT in the test topology shown in Figure 5 in [[I-D.sip-bench-term](#)]. If the DUT does not support loop detection, then this step can be skipped.
2. Configure a SUT according to the test topology shown in Figure 8 of [[I-D.sip-bench-term](#)].
3. Set media streams per session to 0.
4. Turn on the Loop Detection option in the DUT or SUT.
5. Execute benchmarking algorithm as defined in [Section 4.9](#) to get the session establishment rate with loop detection enabled. This rate MUST be recorded using any pertinent parameters as shown in the reporting format of [Section 5.1](#).

**Expected Results:** Session Establishment Rate results obtained with Loop Detection may be lower than those obtained without Loop Detection enabled.

### **6.6. Session Establishment Rate with Forking**

**Objective:**

To benchmark the Session Establishment Rate of the DUT/SUT with zero failures when the Forking Option is enabled.

**Procedure:**

1. If the DUT is being benchmarked as a proxy or B2BUA, and forking is supported in the DUT, then configure the DUT in the test topology shown in Figure 5 in [[I-D.sip-bench-term](#)]. If the DUT does not support forking, then this step can be skipped.
2. Configure a SUT according to the test topology shown in Figure 8 of [[I-D.sip-bench-term](#)].
3. Set media streams per session to 0.
4. Set the number of endpoints that will receive the forked invitation to a value of 2 or more (subsequent tests may increase this value at the discretion of the tester.)
5. Execute benchmarking algorithm as defined in [Section 4.9](#) to get the session establishment rate with forking. This rate MUST be recorded using any pertinent parameters as shown in the reporting format of [Section 5.1](#).





Expected Results: Session Establishment Rate results obtained with Forking may be lower than those obtained without Forking enabled.

#### **6.7. Session Establishment Rate with Forking and Loop Detection**

**Objective:**

To benchmark the Session Establishment Rate of the DUT/SUT with zero failures when both the Forking and Loop Detection Options are enabled.

**Procedure:**

1. If the DUT is being benchmarked as a proxy or B2BUA, then configure the DUT in the test topology shown in Figure 5 in [\[I-D.sip-bench-term\]](#).
2. Configure a SUT according to the test topology shown in Figure 8 of [\[I-D.sip-bench-term\]](#).
3. Set media streams per session to 0.
4. Enable the Loop Detection Options on the DUT.
5. Set the number of endpoints that will receive the forked invitation to a value of 2 or more (subsequent tests may increase this value at the discretion of the tester.)
6. Execute benchmarking algorithm as defined in [Section 4.9](#) to get the session establishment rate with forking and loop detection. This rate MUST be recorded using any pertinent parameters as shown in the reporting format of [Section 5.1](#).

Expected Results: Session Establishment Rate results obtained with Forking and Loop Detection may be lower than those obtained with only Forking or Loop Detection enabled.

#### **6.8. Session Establishment Rate with TLS Encrypted SIP**

**Objective:**

To benchmark the Session Establishment Rate of the DUT/SUT with zero failures when using TLS encrypted SIP.

**Procedure:**

1. If the DUT is being benchmarked as a proxy or B2BUA, then configure the DUT in the test topology shown in Figure 5 in [\[I-D.sip-bench-term\]](#).
2. Configure a SUT according to the test topology shown in Figure 8 of [\[I-D.sip-bench-term\]](#).
3. Set media streams per session to 0.
4. Configure Tester to enable TLS over the transport being benchmarked. Make a note the transport when compiling results. May need to run for each transport of interest.



5. Execute benchmarking algorithm as defined in [Section 4.9](#) to get the session establishment rate with encryption. This rate MUST be recorded using any pertinent parameters as shown in the reporting format of [Section 5.1](#).

Expected Results: Session Establishment Rate results obtained with TLS Encrypted SIP may be lower than those obtained with plaintext SIP.

#### **6.9. Session Establishment Rate with IPsec Encrypted SIP**

**Objective:**

To benchmark the Session Establishment Rate of the DUT/SUT with zero failures when using IPsec Encrypted SIP.

**Procedure:**

1. If the DUT is being benchmarked as a proxy or B2BUA, then configure the DUT in the test topology shown in Figure 5 in [\[I-D.sip-bench-term\]](#).
2. Configure a SUT according to the test topology shown in Figure 8 of [\[I-D.sip-bench-term\]](#).
3. Set media streams per session to 0.
4. Configure Tester for IPsec.
5. Execute benchmarking algorithm as defined in [Section 4.9](#) to get the session establishment rate with encryption. This rate MUST be recorded using any pertinent parameters as shown in the reporting format of [Section 5.1](#).

Expected Results: Session Establishment Rate results obtained with IPsec Encrypted SIP may be lower than those obtained with plaintext SIP.

#### **6.10. Session Establishment Rate with SIP Flooding**

**Objective:**

To benchmark the Session Establishment Rate of the SUT with zero failures when SIP Flooding is occurring.

**Procedure:**

1. If the DUT is being benchmarked as a proxy or B2BUA, then configure the DUT in the test topology shown in Figure 5 in [\[I-D.sip-bench-term\]](#).
2. Configure a SUT according to the test topology shown in Figure 8 of [\[I-D.sip-bench-term\]](#).
3. Set media streams per session to 0.
4. Set  $s = 500$  (c.f. [Section 4.9](#)).



5. Execute benchmarking algorithm as defined in [Section 4.9](#) to get the session establishment rate with flooding. This rate MUST be recorded using any pertinent parameters as shown in the reporting format of [Section 5.1](#).

Expected Results: Session Establishment Rate results obtained with SIP Flooding may be degraded.

#### **[6.11](#). Maximum Registration Rate**

Objective:

To benchmark the maximum registration rate of the DUT/SUT with zero failures.

Procedure:

1. If the DUT is being benchmarked as a proxy or B2BUA, then configure the DUT in the test topology shown in Figure 5 in [\[I-D.sip-bench-term\]](#).
2. Configure a SUT according to the test topology shown in Figure 8 of [\[I-D.sip-bench-term\]](#).
3. Set media streams per session to 0.
4. Set the registration timeout value to at least 3600 seconds.
5. Execute benchmarking algorithm as defined in [Section 4.9](#) to get the maximum registration rate. This rate MUST be recorded using any pertinent parameters as shown in the reporting format of [Section 5.1](#).

Expected Results:

#### **[6.12](#). Maximum Re-Registration Rate**

Objective:

To benchmark the maximum re-registration rate of the DUT/SUT with zero failures.

Procedure:

1. If the DUT is being benchmarked as a proxy or B2BUA, then configure the DUT in the test topology shown in Figure 5 in [\[I-D.sip-bench-term\]](#).
2. Configure a SUT according to the test topology shown in Figure 8 of [\[I-D.sip-bench-term\]](#).
3. First, execute test detailed in [Section 6.11](#) to register the endpoints with the registrar.
4. After at least 5 minutes of Step 2, but no more than 10 minutes after Step 2 has been performed, execute test detailed in [Section 6.11](#) again (this will count as a re-registration).
5. Execute benchmarking algorithm as defined in [Section 4.9](#) to get the maximum re-registration rate. This rate MUST be recorded using any pertinent parameters as shown in the



reporting format of [Section 5.1](#).

Expected Results: The rate should be at least equal to but not more than the result of [Section 6.11](#).

### **6.13. Maximum IM Rate**

Objective:

To benchmark the maximum IM rate of the SUT with zero failures.

Procedure:

1. If the DUT/SUT is being benchmarked as a user agent client or a user agent server, configure the DUT in the test topology shown in Figure 1 or Figure 2 in [[I-D.sip-bench-term](#)]. Alternatively, if the DUT is being benchmarked as a proxy or a B2BUA, configure the DUT in the test topology shown in Figure 5 in [[I-D.sip-bench-term](#)].
2. Configure a SUT according to the test topology shown in Figure 5 in [[I-D.sip-bench-term](#)].
3. Execute benchmarking algorithm as defined in [Section 4.9](#) to get the maximum IM rate. This rate MUST be recorded using any pertinent parameters as shown in the reporting format of [Section 5.1](#).

Expected Results:

### **6.14. Session Capacity without Media**

Objective:

To benchmark the Session Capacity of the SUT without Associated Media.

Procedure:

1. If the DUT/SUT is being benchmarked as a user agent client or a user agent server, configure the DUT in the test topology shown in Figure 1 or Figure 2 in [[I-D.sip-bench-term](#)]. Alternatively, if the DUT is being benchmarked as a proxy or a B2BUA, configure the DUT in the test topology shown in Figure 5 in [[I-D.sip-bench-term](#)].
2. Configure a SUT according to the test topology shown in Figure 7 in [[I-D.sip-bench-term](#)].
3. Set the media streams per session to be 0.
4. Set the Session Duration to be a value greater than T.
5. Execute benchmarking algorithm as defined in [Section 4.9](#) to get the baseline session establishment rate. This rate MUST be recorded using any pertinent parameters as shown in the reporting format of [Section 5.1](#).
6. The Session Capacity is the product of T and the Session Establishment Rate.





Expected Results: The maximum rate at which the DUT/SUT can handle session establishment requests with no media for an infinitely long period with no errors. This is the SIP "throughput" of the system with no media.

#### **6.15. Session Capacity with Media**

Objective:

To benchmark the session capacity of the DUT/SUT with Associated Media.

Procedure:

1. Configure the DUT in the test topology shown in Figure 3 or Figure 4 of [[I-D.sip-bench-term](#)] depending on whether the DUT is being benchmarked as a user agent client or user agent server. Alternatively, configure the DUT in the test topology shown in Figure 6 or Figure 7 in [[I-D.sip-bench-term](#)] depending on whether the DUT is being benchmarked as a B2BUA or as a proxy. If a SUT is being benchmarked, configure the SUT as shown in Figure 9 of [[I-D.sip-bench-term](#)].
2. Set the media streams per session to 1.
3. Set the Session Duration to be a value greater than T.
4. Execute benchmarking algorithm as defined in [Section 4.9](#) to get the baseline session establishment rate. This rate MUST be recorded using any pertinent parameters as shown in the reporting format of [Section 5.1](#).
5. The Session Capacity is the product of T and the Session Establishment Rate.

Expected Results: Session Capacity results obtained with Associated Media with any number of media streams per SIP session will be identical to the Session Capacity results obtained without media.

#### **6.16. Session Capacity with Media and a Media Relay/NAT and/or Firewall**

Objective:

To benchmark the Session Establishment Rate of the SUT with Associated Media.

Procedure:

1. Configure the SUT as shown in Figure 7 or Figure 10 in [[I-D.sip-bench-term](#)].
2. Set media streams per session to 1.
3. Execute benchmarking algorithm as defined in [Section 4.9](#) to get the session establishment rate with media. This rate MUST be recorded using any pertinent parameters as shown in the reporting format of [Section 5.1](#).



Expected Results: Session Capacity results obtained with Associated Media with any number of media streams per SIP session may be lower than the Session Capacity without Media result if the Media Relay, NAT or Firewall is sharing a platform with the server.

## **7. IANA Considerations**

This document does not requires any IANA considerations.

## **8. Security Considerations**

Documents of this type do not directly affect the security of Internet or corporate networks as long as benchmarking is not performed on devices or systems connected to production networks. Security threats and how to counter these in SIP and the media layer is discussed in [RFC3261](#), [RFC3550](#), and [RFC3711](#) and various other drafts. This document attempts to formalize a set of common methodology for benchmarking performance of SIP devices in a lab environment.

## **9. Acknowledgments**

The authors would like to thank Keith Drage and Daryl Malas for their contributions to this document. Dale Worley provided an extensive review that lead to improvements in the documents.

## **10. References**

### **10.1. Normative References**

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC2544] Bradner, S. and J. McQuaid, "Benchmarking Methodology for Network Interconnect Devices", [RFC 2544](#), March 1999.
- [I-D.sip-bench-term]  
Davids, C., Gurbani, V., and S. Poretsky, "SIP Performance Benchmarking Terminology",  
[draft-ietf-bmwg-sip-bench-term-04](#) (work in progress),  
March 2012.



## **10.2. Informative References**

[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.

### Authors' Addresses

Carol Davids  
Illinois Institute of Technology  
201 East Loop Road  
Wheaton, IL 60187  
USA

Phone: +1 630 682 6024  
Email: [davids@iit.edu](mailto:davids@iit.edu)

Vijay K. Gurbani  
Bell Laboratories, Alcatel-Lucent  
1960 Lucent Lane  
Rm 9C-533  
Naperville, IL 60566  
USA

Phone: +1 630 224 0216  
Email: [vkg@bell-labs.com](mailto:vkg@bell-labs.com)

Scott Poretsky  
Allot Communications  
300 TradeCenter, Suite 4680  
Woburn, MA 08101  
USA

Phone: +1 508 309 2179  
Email: [sporetsky@allot.com](mailto:sporetsky@allot.com)

