Benchmarking Methodology Working Group C. Davids Internet-Draft Illinois Institute of Technology Intended status: Informational V. Gurbani Bell Laboratories. Expires: November 29, 2014 Alcatel-Lucent S. Poretsky Allot Communications May 28, 2014

Methodology for Benchmarking Session Initiation Protocol (SIP) Devices: Basic session setup and registration draft-ietf-bmwg-sip-bench-meth-10

### Abstract

This document provides a methodology for benchmarking the Session Initiation Protocol (SIP) performance of devices. Terminology related to benchmarking SIP devices is described in the companion terminology document. Using these two documents, benchmarks can be obtained and compared for different types of devices such as SIP Proxy Servers, Registrars and Session Border Controllers. The term "performance" in this context means the capacity of the device-undertest (DUT) to process SIP messages. Media streams are used only to study how they impact the signaling behavior. The intent of the two documents is to provide a normalized set of tests that will enable an objective comparison of the capacity of SIP devices. Test setup parameters and a methodology are necessary because SIP allows a wide range of configuration and operational conditions that can influence performance benchmark measurements.

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 November 29, 2014.

Davids, et al. Expires November 29, 2014

[Page 1]

Copyright Notice

Copyright (c) 2014 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

$\underline{1}.  \text{Terminology}  \ldots  \ldots  \ldots  \ldots  \ldots  \ldots  \ldots  \ldots  \ldots  $			4			
<u>2</u> . Introduction						
3. Benchmarking Topologies			5			
$\underline{4}$ . Test Setup Parameters						
<u>4.1</u> . Selection of SIP Transport Protocol			6			
<u>4.2</u> . Signaling Server			<u>6</u>			
<u>4.3</u> . Associated Media			7			
<u>4.4</u> . Selection of Associated Media Protocol			7			
4.5. Number of Associated Media Streams per SIP Session .			7			
4.6. Session Duration			7			
4.7. Attempted Sessions per Second (sps)			7			
4.8. Benchmarking algorithm			7			
5. Reporting Format			10			
5.1. Test Setup Report			10			
5.2. Device Benchmarks for IS			10			
5.3. Device Benchmarks for NS			10			
6. Test Cases						
6.1. Baseline Session Establishment Rate of the test bed			11			
6.2. Session Establishment Rate without media			11			
6.3. Session Establishment Rate with Media not on DUT			11			
6.4. Session Establishment Rate with Media on DUT			12			
6.5. Session Establishment Rate with TLS Encrypted SIP .			12			
6.6. Session Establishment Rate with IPsec Encrypted SIP			13			
6.7. Registration Rate			13			
6.8. Re-Registration Rate			14			
7. IANA Considerations			14			
8. Security Considerations			14			
9. Acknowledgments			14			
10. References						
10.1. Normative References						
10.2. Informative References						
Appendix A. R code to simulate benchmarking algorithm						
Authors' Addresses						

# **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.

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].

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

#### 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. Media streams, when used, are used only to study how they impact the signaling behavior. This document concentrates on benchmarking SIP session setup and SIP registrations only.

The device-under-test (DUT) is a SIP server, which may be any SIP conforming [RFC3261] device. Benchmarks can be obtained and compared for different types of devices such as a SIP proxy server, Session Border Controllers (SBC), SIP registrars and a SIP proxy server paired with a media relay.

The test cases provide metrics for benchmarking the maximum 'SIP Registration Rate' and maximum 'SIP Session Establishment Rate' that the DUT can sustain over an extended period of time without failures (extended period of time is defined in the algorithm in Section 4.8). Some cases are included to cover encrypted SIP. The test topologies that can be used are described in the Test Setup section. Topologies in which the DUT handles media as well as those in which the DUT does not handle media are both considered. The measurement of the performance characteristics of the media itself is outside the scope of these documents.

SIP permits a wide range of configuration options that are explained in Section 4 and Section 2 of [I-D.sip-bench-term]. Benchmark values could possibly be impacted by Associated Media. The selected values

Davids, et al. Expires November 29, 2014

[Page 4]

for Session Duration and Media Streams per Session enable benchmark metrics to be evaluated without Associated Media. Session Establishment 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.

SIP permits a wide range of configuration options that are explained in Section 4 and Section 2 of [I-D.sip-bench-term]. 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

There are two test topologies; one in which the DUT does not process the media (Figure 1) and the other in which it does process media (Figure 2). In both cases, the tester or EA sends traffic into the DUT and absorbs traffic from the DUT. The diagrams in Figure 1 and Figure 2 represent the logical flow of information and do not dictate a particular physical arrangements of the entities.

Test organizations need to be aware that these tests generate large volumes of data and consequently ensure that networking devices like hubs, switches or routers are able to handle the generated volume.

Figure 1 depicts a layout in which the DUT is an intermediary between the two interfaces of the EA. If the test case requires the exchange of media, the media does not flow through the DUT but rather passes directly between the two endpoints. Figure 2 shows the DUT as an intermediary between the two interfaces of the EA. If the test case requires the exchange of media, the media flows through the DUT between the endpoints.

Davids, et al. Expires November 29, 2014

[Page 5]



Figure 1: DUT as an intermediary, end-to-end media



Figure 2: DUT as an intermediary forwarding media

#### **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, TCP, UDP, TLS and websockets. 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). The Signaling Server is a DUT.

Davids, et al. Expires November 29, 2014

[Page 6]

# 4.3. Associated Media

Some tests require Associated Media to be present for each SIP session. The test topologies to be used when benchmarking DUT performance for Associated Media are shown in Figure 1 and Figure 2.

#### **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 DUT for voice, a single media stream is used. When benchmarking a DUT 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

The value of the DUT's 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 a successful SIP establishment. Setting this parameter to the value '0' indicates that a BYE will be sent by the EA immediately after the EA receives a 200 OK to the INVITE. Setting this parameter to a time value greater than the duration of the test indicates that a BYE is never sent.

#### **4.7.** Attempted Sessions per Second (sps)

The value of the DUT's performance benchmarks may vary with the Session Attempt Rate offered by the tester. Session Attempt Rate MUST be reported with the benchmarking results.

## 4.8. Benchmarking algorithm

In order to benchmark the test cases uniformly in Section 6, the algorithm described in this section should be used. A prosaic description of the algorithm and a pseudo-code description are provided below, and a simulation written in the R statistical language is provided in Appendix A.

The goal is to find the largest value, R, a SIP Session Attempt Rate,

Davids, et al. Expires November 29, 2014

[Page 7]

measured in sessions-per-second (sps), which the DUT can process with zero errors over a defined, extended period. This period is defined as the amount of time needed to attempt N SIP sessions, where N is a parameter of test, at the attempt rate, R. An iterative process is used to find this rate. The algorithm corresponding to this process converges to R.

If the DUT vendor provides a value for R, the tester can use this value. Alternatively, in cases where the DUT vendor does not provide a value for R, or in cases where the tester wants to ascertain a vendor provided value using local media characteristics, the algorithm could be run by setting "r = R" and observing the value at convergence.

The algorithm proceeds by initializing "r = 100"; "r" is the session attempt rate. The algorithm dynamically increases and decreases "r" as it converges to the a maximum sps value for R. The dynamic increase and decrease rate is controlled by the weights "w" and "d", respectively. If the DUT vendor provides a value for R, the tester can use that value; however, because the requirements and media characteristics are a function of the test environment, it is best that the tester reflect these requirements during testing and allow the algorithm to converge to R.

The pseudo-code corresponding to the description above follows, and a simulation written in the R statistical language is provided in Appendix A.

;	Param	eters of test, adjust as needed				
Ν	:= 50000	; Global maximum; once largest session rate has				
		; been established, send this many requests before				
		; calling the test a success				
m	$:= \{ \dots \}$	; Other attributes that affect testing, such				
		; as media streams, etc.				
r	:= 100	; Initial session attempt rate (in sessions/sec).				
		; Adjust as needed (for example, if DUT can handle				
		; thousands of calls in steady state, set to				
		; appropriate value in the thousands).				
W	:= 0.10	; Traffic increase weight (0 < w <= 1.0)				
d	:= max(0.	10, w / 2) ; Traffic decrease weight				
;	End o	f parameters of test				
pr	oc find_R					
	R = max_sps(r, m, N) ; Setup r sps, each with m media					
	; charact	eristics until N sessions have been set up.				

```
; Note that if a DUT vendor provides this number, the tester
   ; can use the number as a Session Attempt Rate, R, instead
   ; of invoking max_sps()
end proc
; Iterative process to figure out the largest number of
; sps that we can achieve in order to setup n sessions.
; This function converges to R, the Session Attempt Rate.
proc max_sps(r, m, n)
   s := 0 ; session setup rate
  old_r := 0 ; old session setup rate
  h := 0 ; Return value, R
  count := 0
   ; Note that if w is small (say, 0.10) and r is small
   ; (say, <= 9), the algorithm will not converge since it
   ; uses floor() to increment r dynamically. It is best
   ; off to start with the defaults (w = 0.10 and
   ; r >= 10)
  while (TRUE) {
     s := send_traffic(r, m, n) ; Send r sps, with m media
      ; characteristics until n sessions established.
     if (s == n) {
         if (r > old_r) {
             old_r = r
         }
         else {
             count = count + 1
             if (count >= 10) {
                 # We've converged.
                 h := max(r, old_r)
                 break
             }
         }
         r := floor(r + (w * r))
      }
     else {
         r := floor(r - (d * r))
         d := max(0.10, d / 2)
         W := max(0.10, W / 2)
      }
   }
   return h
end proc
```

[Page 9]

# 5. Reporting Format

# 5.1. Test Setup Report

SIP Transport Protocol = \_\_\_\_\_ (valid values: TCP|UDP|TLS|SCTP|websockets|specify-other) Session Attempt Rate = \_\_\_\_\_ (session 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) Establishment Threshold time = \_\_\_\_\_ (seconds) TLS ciphersuite used (for tests involving TLS) = \_\_\_\_\_ (e.g., TLS\_RSA\_WITH\_AES\_128\_CBC\_SHA) IPSec profile used (for tests involving IPSEC) = \_\_\_\_\_

# 5.2. Device Benchmarks for IS

Session Establishment Rate = \_\_\_\_\_ (sessions per second) Is DUT acting as a media relay (yes/no) = \_\_\_\_\_

# 5.3. Device Benchmarks for NS

Registration Rate = \_\_\_\_ (registrations per second) Re-registration Rate = \_\_\_\_ (registrations 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.
- 2. Set media streams per session to 0.
- 3. Execute benchmarking algorithm as defined in Section 4.8 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 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 on the test bed in question.

### 6.2. Session Establishment Rate without media

Objective:

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

Procedure:

- 1. Configure a DUT according to the test topology shown in Figure 1 or Figure 2.
- 2. Set media streams per session to 0.
- 3. Execute benchmarking algorithm as defined in Section 4.8 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: Find the Session Establishment Rate of the DUT when the EA is not sending media streams.

### 6.3. Session Establishment Rate with Media not on DUT

### Objective:

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

#### Procedure:

- 1. Configure a DUT according to the test topology shown in Figure 1.
- 2. Set media streams per session to 1.
- 3. Execute benchmarking algorithm as defined in Section 4.8 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 DUT is running on a platform separate from the Media Relay.

# 6.4. Session Establishment Rate with Media on DUT

#### Objective:

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

#### Procedure:

- 1. Configure a DUT according to the test topology shown in Figure 2.
- 2. Set media streams per session to 1.
- 3. Execute benchmarking algorithm as defined in Section 4.8 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 may be lower than those obtained without media in the case where the DUT and the Media Relay are running on the same platform.

### 6.5. Session Establishment Rate with TLS Encrypted SIP

## Objective:

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

Procedure:

If the DUT is being benchmarked as a proxy or B2BUA, then 1. configure the DUT in the test topology shown in Figure 1 or Figure 2.

- 2. Configure the tester to enable TLS over the transport being used during benchmarking. Note the ciphersuite being used for TLS and record it in Section 5.1.
- 3. Set media streams per session to 0 (media is not used in this test).
- 4. Execute benchmarking algorithm as defined in Section 4.8 to get the session establishment rate with TLS encryption.

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

## 6.6. Session Establishment Rate with IPsec Encrypted SIP

Objective:

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

Procedure:

- 1. Configure a DUT according to the test topology shown in Figure 1 or Figure 2.
- 2. Set media streams per session to 0 (media is not used in this test).
- 3. Configure tester for IPSec. Note the IPSec profile being used for and record it in Section 5.1.
- 4. Execute benchmarking algorithm as defined in Section 4.8 to get the session establishment rate with encryption.

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

# 6.7. Registration Rate

Objective:

To benchmark the maximum registration rate the DUT can handle over an extended time period with zero failures.

Procedure:

- 1. Configure a DUT according to the test topology shown in Figure 1 or Figure 2.
- 2. Set the registration timeout value to at least 3600 seconds.
- Execute benchmarking algorithm as defined in Section 4.8 to 3. get the maximum registration rate. This rate MUST be recorded using any pertinent parameters as shown in the reporting format of Section 5.1. For example, the use of TLS or IPSec during registration must be noted in the reporting format.

Davids, et al. Expires November 29, 2014

Expected Results: Provides a maximum registration rate.

# 6.8. Re-Registration Rate

### Objective:

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

### Procedure:

- 1. Configure a DUT according to the test topology shown in Figure 1 or Figure 2.
- First, execute test detailed in Section 6.7 to register the 2. endpoints with the registrar and obtain the registration rate.
- 3. After at least 5 minutes of Step 2, but no more than 10 minutes after Step 2 has been performed, execute Step 3 of the test in Section 6.7. This will count as a re-registration because the SIP address of records have not yet expired.

Expected Results: The rate should be at least equal to but not more than the result of Section 6.7.

### 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. We are grateful to Barry Constantine for providing valuable comments during the document's WGLC.

Davids, et al. Expires November 29, 2014

[Page 14]

Internet-Draft SIP Benchmarking Methodology

[Page 15]

# **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-10 (work in progress),
May 2014.
```

# **<u>10.2</u>**. Informative References

[RFC3261] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", <u>RFC 3261</u>, June 2002.

Appendix A. R code to simulate benchmarking algorithm

```
W = 0.10
d = max(0.10, w / 2)
DUT_max_sps = 460  # Change as needed to set the max sps value
                      # for a DUT
# Returns R, given r (initial session attempt rate).
# E.g., assume that a DUT handles 460 sps in steady state
# and you have saved this code in a file simulate.r. Then,
# start an R session and do the following:
#
# > source("simulate.r")
\# > find_R(100)
# ... debug output omitted ...
# [1] 458
#
# Thus, the max sps that the DUT can handle is 458 sps, which is
# close to the absolute maximum of 460 sps the DUT is specified to
# do.
find_R <- function(r) {</pre>
  s = 0
   old_r = 0
```

```
h = 0
  count = 0
  # Note that if w is small (say, 0.10) and r is small
  # (say, <= 9), the algorithm will not converge since it</pre>
  # uses floor() to increment r dynamically. It is best
  # off to start with the defaults (w = 0.10 and
  # r >= 10)
  cat("r old_r w d \n")
  while (TRUE) {
     cat(r, '', old_r, '', w, '', d, '\n')
     s = send_traffic(r)
     if (s == TRUE) { # All sessions succeeded
         if (r > old_r) {
            old_r = r
         }
         else {
             count = count + 1
             if (count >= 10) {
                 # We've converged.
                 h = max(r, old_r)
                 break
             }
         }
         r = floor(r + (w * r))
     }
     else {
         r = floor(r - (d * r))
         d = max(0.10, d / 2)
         W = max(0.10, W / 2)
     }
  }
  h
}
send_traffic <- function(r) {</pre>
  n = TRUE
  if (r > DUT_max_sps) {
      n = FALSE
  }
  n
```

[Page 16]

}

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 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 Scott Poretsky Allot Communications 300 TradeCenter, Suite 4680 Woburn, MA 08101 USA Phone: +1 508 309 2179 Email: sporetsky@allot.com