Benchmarking Methodology Working C. Davids

Group Illinois Institute of Technology

Internet-Draft V. Gurbani

Expires: May 12, 2013 Bell Laboratories, Alcatel-Lucent

 S. Poretsky

 Allot Communications

 November 8, 2012

 Terminology for Benchmarking Session Initiation Protocol (SIP)

 Networking Devices

 draft-ietf-bmwg-sip-bench-term-06

Abstract

 This document provides a terminology for benchmarking the SIP

 performance of networking devices. The term performance in this

 context means the capacity of the device- or system-under-test to

 process SIP messages. Terms are included for test components, test

 setup parameters, and performance benchmark metrics for black-box

 benchmarking of SIP networking devices. The performance benchmark

 metrics are obtained for the SIP signaling plane only. The terms are

 intended for use in a companion methodology document for

 characterizing the performance of a SIP networking device under a

 variety of conditions. The intent of the two documents is to enable

 a comparison of the capacity of SIP networking devices. Test setup

 parameters and a methodology document are necessary because SIP

 allows a wide range of configuration and operational conditions that

 can influence performance benchmark measurements. A standard

 terminology and methodology will ensure that benchmarks have

 consistent definition and were obtained following the same

 procedures.

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."

Davids, et al. Expires May 12, 2013 [Page 1]

Internet-Draft SIP Benchmarking Terminology November 2012

 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.

Davids, et al. Expires May 12, 2013 [Page 2]

Internet-Draft SIP Benchmarking Terminology November 2012

Table of Contents

 1. Terminology . . . . . . . . . . . . . . . . . . . . . . . . . 5

 2. Introduction . . . . . . . . . . . . . . . . . . . . . . . . . 6

 2.1. Scope . . . . . . . . . . . . . . . . . . . . . . . . . . 7

 2.2. Benchmarking Models . . . . . . . . . . . . . . . . . . . 9

 3. Term Definitions . . . . . . . . . . . . . . . . . . . . . . . 14

 3.1. Protocol Components . . . . . . . . . . . . . . . . . . . 14

 3.1.1. Session . . . . . . . . . . . . . . . . . . . . . . . 14

 3.1.2. Signaling Plane . . . . . . . . . . . . . . . . . . . 17

 3.1.3. Media Plane . . . . . . . . . . . . . . . . . . . . . 18

 3.1.4. Associated Media . . . . . . . . . . . . . . . . . . . 18

 3.1.5. Overload . . . . . . . . . . . . . . . . . . . . . . . 19

 3.1.6. Session Attempt . . . . . . . . . . . . . . . . . . . 20

 3.1.7. Established Session . . . . . . . . . . . . . . . . . 20

 3.1.8. Invite-initiated Session (IS) . . . . . . . . . . . . 21

 3.1.9. Non-INVITE-initiated Session (NS) . . . . . . . . . . 22

 3.1.10. Session Attempt Failure . . . . . . . . . . . . . . . 22

 3.1.11. Standing Sessions Count . . . . . . . . . . . . . . . 23

 3.2. Test Components . . . . . . . . . . . . . . . . . . . . . 23

 3.2.1. Emulated Agent . . . . . . . . . . . . . . . . . . . . 24

 3.2.2. Signaling Server . . . . . . . . . . . . . . . . . . . 24

 3.2.3. SIP-Aware Stateful Firewall . . . . . . . . . . . . . 24

 3.2.4. SIP Transport Protocol . . . . . . . . . . . . . . . . 25

 3.3. Test Setup Parameters . . . . . . . . . . . . . . . . . . 26

 3.3.1. Session Attempt Rate . . . . . . . . . . . . . . . . . 26

 3.3.2. IS Media Attempt Rate . . . . . . . . . . . . . . . . 26

 3.3.3. Establishment Threshold Time . . . . . . . . . . . . . 27

 3.3.4. Session Duration . . . . . . . . . . . . . . . . . . . 27

 3.3.5. Media Packet Size . . . . . . . . . . . . . . . . . . 28

 3.3.6. Media Offered Load . . . . . . . . . . . . . . . . . . 28

 3.3.7. Media Session Hold Time . . . . . . . . . . . . . . . 29

 3.3.8. Loop Detection Option . . . . . . . . . . . . . . . . 29

 3.3.9. Forking Option . . . . . . . . . . . . . . . . . . . . 30

 3.4. Benchmarks . . . . . . . . . . . . . . . . . . . . . . . . 31

 3.4.1. Registration Rate . . . . . . . . . . . . . . . . . . 31

 3.4.2. Session Establishment Rate . . . . . . . . . . . . . . 31

 3.4.3. Session Capacity . . . . . . . . . . . . . . . . . . . 32

 3.4.4. Session Overload Capacity . . . . . . . . . . . . . . 33

 3.4.5. Session Establishment Performance . . . . . . . . . . 33

 3.4.6. Session Attempt Delay . . . . . . . . . . . . . . . . 34

 3.4.7. IM Rate . . . . . . . . . . . . . . . . . . . . . . . 34

 4. IANA Considerations . . . . . . . . . . . . . . . . . . . . . 35

 5. Security Considerations . . . . . . . . . . . . . . . . . . . 35

 6. Acknowledgments . . . . . . . . . . . . . . . . . . . . . . . 36

 7. References . . . . . . . . . . . . . . . . . . . . . . . . . . 36

 7.1. Normative References . . . . . . . . . . . . . . . . . . . 36

 7.2. Informational References . . . . . . . . . . . . . . . . . 36

Davids, et al. Expires May 12, 2013 [Page 3]

Internet-Draft SIP Benchmarking Terminology November 2012

 Appendix A. White Box Benchmarking Terminology . . . . . . . . . 37

 Authors' Addresses . . . . . . . . . . . . . . . . . . . . . . . . 37

Davids, et al. Expires May 12, 2013 [Page 4]

Internet-Draft SIP Benchmarking Terminology November 2012

1. Terminology

 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 BCP 14, RFC2119

 [RFC2119]. 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 [RFC2544].

 For the sake of clarity and continuity, this document adopts the

 template for definitions set out in Section 2 of RFC 1242 [RFC1242].

 The terms Device Under Test (DUT) and System Under Test (SUT) are

 defined in the following BMWG documents:

 Device Under Test (DUT) (c.f., Section 3.1.1 RFC 2285 [RFC2285]).

 System Under Test (SUT) (c.f., Section 3.1.2, RFC 2285 [RFC2285]).

 Many commonly used SIP terms in this document are defined in RFC 3261

 [RFC3261]. For convenience the most important of these are

 reproduced below. Use of these terms in this document is consistent

 with their corresponding definition in [RFC3261].

 o Call Stateful: A proxy is call stateful if it retains state for a

 dialog from the initiating INVITE to the terminating BYE request.

 A call stateful proxy is always transaction stateful (should transaction stateful be defined? Seems to be the same as call stateful?), but the

 converse is not necessarily true.

 o Stateful Proxy: A logical entity that maintains the client and

 server transaction state machines defined by this specification

 during the processing of a request, also known as a transaction

 stateful proxy. The behavior of a stateful proxy is further

 defined in Section 16 of RFC 3261 [RFC3261] . A transaction

 stateful proxy is not the same as a call stateful proxy (this is to the point above that transaction Stateful should probably be a separate bullet with definition).

 o Stateless Proxy: A logical entity that does not maintain the

 client or server transaction state machines defined in this

 specification when it processes requests. A stateless proxy

 forwards every request it receives downstream and every response

 it receives upstream.

 o Back-to-back User Agent: A back-to-back user agent (B2BUA) is a

 logical entity that receives a request and processes it as a user

 agent server (UAS). In order to determine how the request should

 be answered, it acts as a user agent client (UAC) and generates

 requests. Unlike a proxy server, it maintains dialog state and

 must participate in all requests sent on the dialogs it has

 established. Since it is a concatenation of a UAC and a UAS, no

 explicit definitions are needed for its behavior.

Davids, et al. Expires May 12, 2013 [Page 5]

Internet-Draft SIP Benchmarking Terminology November 2012

 o Loop: A request that arrives at a proxy, is forwarded, and later

 arrives back at the same proxy. When it arrives the second time,

 its Request-URI is identical to the first time, and other header

 fields that affect proxy operation are unchanged, so that the

 proxy will make the same processing decision on the request it

 made the first time. Looped requests are errors, and the

 procedures for detecting them and handling them are described by

 the SIP protocol[RFC3261] and also by RFC 5393

2. Introduction

 Service Providers and IT Organizations deliver Voice Over IP (VoIP)

 and Multimedia network services based on the IETF Session Initiation

 Protocol (SIP) [RFC3261]. SIP is a signaling protocol originally

 intended to dynamically establish, disconnect and modify

 streams of media between end users. As it has evolved, it has been

 adopted for use in a growing number of services and applications.

 Many of these result in the creation of a media session, but some do

 not. Examples of this latter group include text messaging and

 subscription services. The set of benchmarking terms provided in

 this document is intended for use with any SIP-enabled device

 performing SIP functions in the interior of the network, whether or

 not these result in the creation of media sessions. The performance

 of end-user devices is outside the scope of this document.

 A number of networking devices have been developed to support SIP-

 based VoIP services. These include SIP Servers, Session Border

 Controllers (SBC), Back-to-back User Agents (B2BUA), and SIP-Aware

 Stateful Firewalls. These devices contain a mix of voice and IP

 functions whose performance may be reported using metrics defined by

 the equipment manufacturer or vendor. The Service Provider or IT

 Organization seeking to compare the performance of such devices will

 not be able to do so using these vendor-specific metrics, whose

 conditions of test and algorithms for collection are often

 unspecified. SIP functional elements and the devices that include

 them can be configured many different ways and can be organized into

 various topologies. These configuration and topological choices

 impact the value of any chosen signaling benchmark. Unless these

 conditions-of-test are defined, a true comparison of performance

 metrics will not be possible. Some SIP-enabled network devices

 terminate or relay media as well as signaling. The processing of

 media by the device impacts the signaling performance. As a result,

 the conditions-of-test must include information as to whether or not

 the device under test processes media and if the device does process

 media, a description of the media handled and the manner in which it

 is handled (check to make sure this is described in later sections, by itself it is not clear). This document and its companion methodology document

 [I-D.ietf-bmwg-sip-bench-meth] provide a set of black-box benchmarks

Davids, et al. Expires May 12, 2013 [Page 6]

Internet-Draft SIP Benchmarking Terminology November 2012

 for describing and comparing the performance of devices that

 incorporate the SIP User Agent Client and Server functions and that

 operate in the network's core.

 The definition of SIP performance benchmarks necessarily includes

 definitions of Test Setup Parameters and a test methodology. These

 enable the Tester to perform benchmarking tests on different devices

 and to achieve comparable results. This document provides a common

 set of definitions for Test Components, Test Setup Parameters, and

 Benchmarks. All the benchmarks defined are black-box measurements of

 the SIP signaling plane. The Test Setup Parameters and Benchmarks

 defined in this document are intended for use with the companion

 Methodology document. Benchmarks of internal DUT characteristics

 (also known as white-box benchmarks) such as Session Attempt Arrival

 Rate, which is measured at the DUT, are described in Appendix A to

 allow additional characterization of DUT behavior with different

 distribution models.

2.1. Scope

 The scope of this work item is summarized as follows:

 o This terminology document describes SIP signaling performance

 benchmarks for black-box measurements of SIP networking devices.

 Stress and debug scenarios are not addressed in this work item.

 o The DUT must be an RFC 3261 capable network equipment. This may

 be a Registrar, Redirect Server, Stateless Proxy or Stateful

 Proxy. A DUT MAY also include a B2BUA, SBC functionality. The

 DUT MAY be a multi-port SIP-to-switched network gateway

 implemented as a SIP UAC or UAS.

 o The DUT MAY include an internal SIP Application Level Gateway

 (ALG), firewall, and/or a Network Address Translator (NAT). This

 is referred to as the "SIP Aware Stateful Firewall."

 o The DUT or SUT MUST NOT be end user equipment, such as personal

 digital assistant, a computer-based client, or a user terminal.

 o The Tester acts as multiple "Emulated Agents" (EA) that initiate

 (or respond to) SIP messages as session endpoints and source (or

 receive) associated media for established connections.

 o SIP Signaling in presence of Media

 \* The media performance is not benchmarked in this work item.

 \* It is RECOMMENDED that SIP signaling plane benchmarks be

 performed with media present, but this is optional.

 \* The SIP INVITE requests MUST include the SDP body.

 \* The type of DUT dictates whether the associated media streams

 traverse the DUT or SUT. Both scenarios are within the scope

 of this work item.

 \* SIP is frequently used to create media streams; the signaling

 plane and media plane are treated as orthogonal to each other

 in this document. While many devices support the creation of

Davids, et al. Expires May 12, 2013 [Page 7]

Internet-Draft SIP Benchmarking Terminology November 2012

 media streams, benchmarks that measure the performance of these

 streams are outside the scope of this document and its

 companion methodology document [I-D.ietf-bmwg-sip-bench-meth].

 Tests may be performed with or without the creation of media

 streams. The presence or absence of media streams MUST be

 noted as a condition of the test as the performance of SIP

 devices may vary accordingly. Even if the media is used during

 benchmarking, only the SIP performance will be benchmarked, not

 the media performance or quality.

 o Both INVITE and non-INVITE scenarios (such as Instant Messages or

 IM) are addressed in this document. However, benchmarking SIP

 presence is not a part of this work item.

 o Different transport mechanisms -- such as UDP, TCP, SCTP, or TLS

 -- may be used. The specific transport mechanism MUST be noted as

 a condition of the test as the performance of SIP devices may vary

 accordingly.

 o Looping and forking options are also considered since they impact

 processing at SIP proxies.

 o REGISTER and INVITE requests may be challenged or remain

 unchallenged for authentication purpose. Whether or not the

 REGISTER and INVITE requests are challenged is a condition of test

 which will be recorded along with other such parameters which may

 impact the SIP performance of the device or system under test.

 o Re-INVITE requests are not considered in scope of this work item

 since the benchmarks for INVITEs are based on the dialog created

 by the INVITE and not on the transactions that take place within

 that dialog.

 o Only session establishment is considered for the performance

 benchmarks. Session disconnect is not considered in the scope of

 this work item. This is because our goal is to determine the

 maximum throughput of the device or system under test, that is the

 number of simultaneous SIP sessions that the device or system can

 support. It is true that there are BYE requests being created

 during the test process. These transactions do contribute to the

 load on the device or system under test and thus are accounted for

 in the metric we derive. We do not seek a separate metric for the

 number of BYE transactions a device or system can support.

 o SIP Overload [I-D.ietf-soc-overload-design] is within the scope of

 this work item. We test to failure and then can continue to

 observe and record the behavior of the system after failures are

 recorded. The cause of failure is not within the scope of this

 work. (I want to understand this more in the methodology document, seems like a subjective area as far as benchmarking. The recovery time seems more consistent with benchmarking similar to system recovery benchmark in RFC 2544). We note the failure and may continue to test until a

 different failure or condition is encountered. Considerations on

 how to handle overload are deferred to work progressing in the SOC

 working group [I-D.ietf-soc-overload-control]. Vendors are, of

 course, free to implement their specific overload control behavior

 as the expected test outcome if it is different from the IETF

 recommendations. However, such behavior MUST be documented and

Davids, et al. Expires May 12, 2013 [Page 8]

Internet-Draft SIP Benchmarking Terminology November 2012

 interpreted appropriately across multiple vendor implementations.

 This will make it more meaningful to compare the performance of

 different SIP overload implementations.

 o IMS-specific scenarios are not considered, but test cases can be

 applied with 3GPP-specific SIP signaling and the P-CSCF as a DUT.

2.2. Benchmarking Models

 This section shows ten models to be used when benchmarking SIP

 performance of a networking device. Figure 1 shows the

 configuration needed to benchmark the tester itself. This model will

 be used to establish the limitations of the test apparatus.

 +--------+ Signaling request +--------+

 | +----------------------------->| |

 | Tester | | Tester |

 | EA | Signaling response | EA |

 | |<-----------------------------+ |

 +--------+ +--------+

 /|\ /|\

 | Media |

 +=========================================+

 Figure 1: Baseline performance of the Emulated Agent without a DUT

 present

 Figure 2 shows the DUT playing the role of a user agent client (UAC),

 initiating requests and absorbing responses (thought this work was not going to benchmark UAC?). This model can be used

 to baseline the performance of the DUT acting as an UAC without

 associated media.

 +--------+ Signaling request +--------+

 | +----------------------------->| |

 | DUT | | Tester |

 | | Signaling response | EA |

 | |<-----------------------------+ |

 +--------+ +--------+

 Figure 2: Baseline performance for DUT acting as a user agent client

 without associated media

 Figure 3 shows the DUT playing the role of a user agent server (UAS),

 absorbing the requests and sending responses. This model can be used

 as a baseline performance for the DUT acting as a UAS without

Davids, et al. Expires May 12, 2013 [Page 9]

Internet-Draft SIP Benchmarking Terminology November 2012

 associated media.

 +--------+ Signaling request +--------+

 | +----------------------------->| |

 | Tester | | DUT |

 | EA | Response | |

 | |<-----------------------------+ |

 +--------+ +--------+

 Figure 3: Baseline performance for DUT acting as a user agent server

 without associated media

 Figure 4 shows the DUT plays the role of a user agent client (UAC),

 initiating requests and absorbing responses (same comments as Figure 2). This model can be used

 as a baseline performance for the DUT acting as a UAC with associated

 media.

 +--------+ Signaling request +--------+

 | +----------------------------->| |

 | DUT | | Tester |

 | | Signaling response | (EA) |

 | |<-----------------------------+ |

 | |<============ Media =========>| |

 +--------+ +--------+

 Figure 4: Baseline performance for DUT acting as a user agent client

 with associated media

 Figure 5 shows the DUT plays the role of a user agent server (UAS),

 absorbing the requests and sending responses. This model can be used

 as a baseline performance for the DUT acting as a UAS with associated

 media.

 +--------+ Signaling request +--------+

 | +----------------------------->| |

 | Tester | | DUT |

 | (EA) | Response | |

 | |<-----------------------------+ |

 | |<============ Media =========>| |

 +--------+ +--------+

 Figure 5: Baseline performance for DUT acting as a user agent server

Davids, et al. Expires May 12, 2013 [Page 10]

Internet-Draft SIP Benchmarking Terminology November 2012

 with associated media

 Figure 6 shows that the Tester acts as the initiating and responding

 EA as the DUT/SUT forwards Session Attempts (and the DUT is acting as Proxy?).

 +--------+ Session +--------+ Session +--------+

 | | Attempt | | Attempt | |

 | |<------------+ |<------------+ |

 | | | | | |

 | | Response | | Response | |

 | Tester +------------>| DUT +------------>| Tester |

 | (EA) | | | | (EA) |

 | | | | | |

 +--------+ +--------+ +--------+

 Figure 6: DUT/SUT performance benchmark for session establishment

 without media

 Figure 7 is used when performing those same benchmarks with

 Associated Media traversing the DUT/SUT.

 +--------+ Session +--------+ Session +--------+

 | | Attempt | | Attempt | |

 | |<------------+ |<------------+ |

 | | | | | |

 | | Response | | Response | |

 | Tester +------------>| DUT +------------>| Tester |

 | (EA) | | | | (EA) |

 | | Media | | Media | |

 | |<===========>| |<===========>| |

 +--------+ +--------+ +--------+

 Figure 7: DUT/SUT performance benchmark for session establishment

 with media traversing the DUT

 Figure 8 is to be used when performing those same benchmarks with

 Associated Media, but the media does not traverse the DUT/SUT.

 Again, the benchmarking of the media is not within the scope of this

 work item. The SIP control signaling is benchmarked in the presence

 of Associated Media to determine if the SDP body of the signaling and

 the handling of media impacts the performance of the DUT/SUT. If the media does not flow through the DUT, then I don’t see the value. This will not affect the SIP signaling performance, right?

Davids, et al. Expires May 12, 2013 [Page 11]

Internet-Draft SIP Benchmarking Terminology November 2012

 +--------+ Session +--------+ Session +--------+

 | | Attempt | | Attempt | |

 | |<------------+ |<------------+ |

 | | | | | |

 | | Response | | Response | |

 | Tester +------------>| DUT +------------>| Tester |

 | (EA) | | | | (EA) |

 | | | | | |

 +--------+ +--------+ +--------+

 /|\ /|\

 | Media |

 +=============================================+

 Figure 8: DUT/SUT performance benchmark for session establishment

 with media external to the DUT

 Figure 9 is used when performing benchmarks that require one or more

 intermediaries to be in the signaling path. The intent is to gather

 benchmarking statistics with a series of DUTs in place. In this

 topology, the media is delivered end-to-end and does not traverse the

 DUT. Are both of the DUTs same vendor / model number ? Seems like it would be complex to isolate the bottleneck unless this was the case. Same comment on media. If the media external can be eliminated, this would simplify the scenarios too which are becoming a little over welming.

 SUT

 ------------------^^^^^^^^-------------

 / \

 +------+ Session +---+ Session +---+ Session +------+

 | | Attempt | | Attempt | | Attempt | |

 | |<---------+ |<---------+ |<---------+ |

 | | | | | | | |

 | | Response | | Response | | Response | |

 |Tester+--------->|DUT+--------->|DUT|--------->|Tester|

 | (EA) | | | | | | (EA) |

 | | | | | | | |

 +------+ +---+ +---+ +------+

 /|\ /|\

 | Media |

 +=============================================+

 Figure 9: DUT/SUT performance benchmark for session establishment

 with multiple DUTs and end-to-end media

 Figure 10 is used when performing benchmarks that require one or more

 intermediaries to be in the signaling path. The intent is to gather

 benchmarking statistics with a series of DUTs in place. In this

 topology, the media is delivered hop-by-hop through each DUT.

Davids, et al. Expires May 12, 2013 [Page 12]

Internet-Draft SIP Benchmarking Terminology November 2012

 SUT

 -----------------^^^^^^^^-------------

 / \

 +------+ Session +---+ Session +---+ Session +------+

 | | Attempt | | Attempt | | Attempt | |

 | |<---------+ |<---------+ |<---------+ |

 | | | | | | | |

 | | Response | | Response | | Response | |

 |Tester+--------->|DUT+--------->|DUT|--------->|Tester|

 | (EA) | | | | | | (EA) |

 | | | | | | | |

 | |<========>| |<========>| |<========>| |

 +------+ Media +---+ Media +---+ Media +------+

 Figure 10: DUT/SUT performance benchmark for session establishment

 with multiple DUTs and hop- by-hop media

 Figure 11 illustrates the SIP signaling for an Established Session.

 The Tester acts as the EAs and initiates a Session Attempt with the

 DUT/SUT. When the EA receives a 200 OK from the

 DUT/SUT, that session is considered to be an Established Session. The

 illustration indicates three states of the session bring created by

 the EA – (1) Attempting, (2) Established, and (3)Disconnecting.

Sessions can be

 one of two type: Invite-Initiated Session (IS) or Non-Invite

 Initiated Session (NS). Failure for the DUT/SUT to successfully

 respond within the Establishment Threshold Time is considered a

 Session Attempt Failure. SIP Invite messages MUST include the SDP

 body to specify the Associated Media. Use of Associated Media, to be

 sourced from the EA, is optional. When Associated Media is used, it

 may traverse the DUT/SUT depending upon the type of DUT/SUT. The

 Associated Media is shown in Figure 11 as "Media" connected to media

 ports M1 and M2 on the EA. After the EA sends a BYE, the session

 disconnects. Performance test cases for session disconnects are not

 considered in this work item (the BYE request is shown for

 completeness.)

Davids, et al. Expires May 12, 2013 [Page 13]

Internet-Draft SIP Benchmarking Terminology November 2012

 EA DUT/SUT M1 M2

 | | | |

 | INVITE | | |

 ---------+-------------->| | |

 | | | |

 Attempting | | |

 | 200 OK | | |

 ---------+<--------------| | |

 | ACK | | |

 |-------------->| | |

 | | | |

 | | | |

 | | | Media |

 Established | |<=====>|

 | | | |

 | BYE | | |

 --------+--------------> | | |

 | | | |

 Disconnecting | | |

 | 200 OK | | |

 --------|<-------------- | | |

 | | | |

 Figure 11: Invite-initiated Session States

3. Term Definitions

3.1. Protocol Components (should the fact that the signaling can occur over TCP or UDP, TCP being more common? Reverse for media.)

3.1.1. Session

 Definition:

 The combination of signaling and media messages and processes that

 support a SIP-based service.

 Discussion:

 SIP messages are used to create and manage services for end users.

 Often, these services include the creation of media streams that

 are defined in the SDP body of a SIP message and carried in RTP

 protocol data units. However, SIP messages can also be used to

 create Instant Message services and subscription services, and

 such services are not associated with media streams. SIP reserves

 the term "session" to describe services that are analogous to

 telephone calls on a circuit switched network. SIP reserves the

 term "dialog" to refer to a signaling-only relationship between

 User Agent peers. SIP reserves the term "transaction" to refer to

Davids, et al. Expires May 12, 2013 [Page 14]

Internet-Draft SIP Benchmarking Terminology November 2012

 the brief communication between a client and a server that lasts

 only until the final response to the SIP request. None of these

 terms describes the entity whose performance we want to benchmark.

 For example, the MESSAGE request does not create a dialog and can

 be sent either within or outside of a dialog. It is not

 associated with media, but it resembles a phone call in its

 dependence on human rather than machine initiated responses. The

 SUBSCRIBE method does create a dialog between the originating end-

 user and the subscription service. It, too, is not associated

 with a media session.

 In light of the above observations, we have extended the term

 "session" to include SIP-based services that are not initiated by

 INVITE requests and that do not have associated media. In this

 extended definition, a session always has a signaling component

 and may also have a media component (this confusing, the above sentence says that the extended session does not have media). Thus, a session can be

 defined as signaling-only or a combination of signaling and media.

 We define the term "Associated Media", see Section 3.1.4, to

 describe the situation in which media is associated with a SIP

 dialog. The terminology "Invite-initiated Session" (IS)

 Section 3.1.8 and "Non-invite-Initiated Session" (NS)

 Section 3.1.9 are used to distinguish between these two types of

 sessions. An Invite-initiated Session is a session as defined in

 SIP. The performance of a device or system that supports Invite-

 initiated Sessions that do not create media sessions, "Invite-

 initiated Sessions without Associated Media", can be measured and

 is of interest for comparison and as a limiting case. The

 REGISTER request can be considered to be a "Non-invite-initiated

 Session without Associated Media." A separate set of benchmarks

 is provided for REGISTER requests since most implementations of

 SIP-based services require this request and since a registrar may

 be a device under test.

 A Session in the context of this document, can be considered to be

 a vector with three components:

 1. A component in the signaling plane (SIP messages), sess.sig;

 2. A media component in the media plane (RTP and SRTP streams for

 example), sess.med (which may be null);

 3. A control component in the media plane (RTCP messages for

 example), sess.medc (which may be null).

 An IS is expected to have non-null sess.sig and sess.med

 components. The use of control protocols in the media component

 is media dependent, thus the expected presence or absence of

 sess.medc is media dependent and test-case dependent. An NS is

 expected to have a non-null sess.sig component, but null sess.med

 and sess.medc components. Is this information really important ? If not, it is hard to follow, guess I’ll see as I read through.

Davids, et al. Expires May 12, 2013 [Page 15]

Internet-Draft SIP Benchmarking Terminology November 2012

 Packets in the Signaling Plane and Media Plane will be handled by

 different processes within the DUT. They will take different

 paths within a SUT. These different processes and paths may

 produce variations in performance. The terminology and benchmarks

 defined in this document and the methodology for their use are

 designed to enable us to compare performance of the DUT/SUT with

 reference to the type of SIP-supported application it is handling.

 Note that one or more sessions can simultaneously exist between

 any participants. This can be the case, for example, when the EA

 sets up both an IM and a voice call through the DUT/SUT. These

 sessions are represented as an array session[x].

 Sessions will be represented as a vector array with three

 components, as follows:

 session->

 session[x].sig, the signaling component

 session[x].medc[y], the media control component (e.g. RTCP)

 session[x].med[y], an array of associated media streams (e.g.

 RTP, SRTP, RTSP, MSRP). This media component may consist of zero

 or more media streams.

 Figure 12 models the vectors of the session. Might be my lack of experience in this vector representation, but I wonder if industry lab guy running these tests will get any value out of these representations.

 Measurement Units:

 N/A.

 Issues:

 None.

 See Also:

 Media Plane

 Signaling Plane

 Associated Media

 Invite-initiated Session (IS)

 Non-invite-initiated Session (NS)

Davids, et al. Expires May 12, 2013 [Page 16]

Internet-Draft SIP Benchmarking Terminology November 2012

 |\

 |

 | \

 sess.sig|

 | \

 |

 | \

 | o

 | /

 | / |

 | /

 | / |

 | /

 | / |

 | /

 | / | sess.medc

 |/\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

 / /

 / |

 / /

 sess.med / |

 /\_ \_ \_ \_ \_ \_ \_ \_/

 /

 /

 /

 /

 Figure 12: Session components

3.1.2. Signaling Plane

 Definition:

 The plane in which SIP messages [RFC3261] are exchanged between

 SIP Agents [RFC3261].

 Discussion:

 SIP messages are used to establish sessions in several ways:

 directly between two User Agents [RFC3261], through a Proxy Server

 [RFC3261], or through a series of Proxy Servers. The Session

 Description Protocol is included in the Signaling Plane. (SDP). This SDP seems dangling.

 The Signaling Plane for a single Session is represented by

 session.sig.

Davids, et al. Expires May 12, 2013 [Page 17]

Internet-Draft SIP Benchmarking Terminology November 2012

 Measurement Units:

 N/A. Why N/A, wouldn’t these be calls per second, etc?

 Issues:

 None.

 See Also:

 Media Plane

 EAs

3.1.3. Media Plane

 Definition:

 The data plane in which one or more media streams and their

 associated media control protocols are exchanged between User

 Agents after a media connection has been created by the exchange

 of signaling messages in the Signaling Plane.

 Discussion:

 Media may also be known as the "bearer channel". The Media Plane

 MUST include the media control protocol, if one is used, and the

 media stream(s). Examples of media are audio and video. The

 media streams are described in the SDP of the Signaling Plane.

 The media for a single Session is represented by session.med. The

 media control protocol for a single media description is

 represented by session.medc.

 Measurement Units:

 N/A.

 Issues:

 None.

 See Also:

 Signaling Plane

3.1.4. Associated Media

 Definition:

 Media that corresponds to an 'm' line in the SDP payload of the

 Signaling Plane.

 Discussion:

Davids, et al. Expires May 12, 2013 [Page 18]

Internet-Draft SIP Benchmarking Terminology November 2012

 Any media protocol MAY be used.

 For any session's signaling component, session.sig, there may be

 zero, one, or multiple associated media streams. When there are

 multiple media streams, these are represented be a vector array

 session.med[y]. When there are multiple media streams there will

 be multiple media control protocol descriptions as well. They are

 represented by a vector array session.medc[y].

 Measurement Units:

 N/A.

 Issues:

 None.

3.1.5. Overload

 Definition:

 Overload is defined as the state where a SIP server does not have

 sufficient resources to process all incoming SIP messages

 [I-D.ietf-soc-overload-design].

 Discussion:

 The distinction between an overload condition and other failure

 scenarios is outside the scope of black box testing and of this

 document. Under overload conditions, all or a percentage of

 Session Attempts will fail due to lack of resources. In black box

 testing the cause of the failure is not explored. The fact that a

 failure occurred for whatever reason, will trigger the tester to

 reduce the offered load, as described in the companion methodology

 document, [I-D.ietf-bmwg-sip-bench-meth]. SIP server resources

 may include CPU processing capacity, network bandwidth, input/

 output queues, or disk resources. Any combination of resources

 may be fully utilized when a SIP server (the DUT/SUT) is in the

 overload condition. For proxy-only type of devices, it is

 expected that the proxy will be driven into overload based on the

 delivery rate of signaling requests.

 For UA-type of network devices such as gateways, it is expected

 that the UA will be driven into overload based on the volume of

 media streams it is processing. This ties to earlier comment, should the device be placed into overload and then benchmarked to determine when system functionality is restored?

 Measurement Units:

 N/A.

Davids, et al. Expires May 12, 2013 [Page 19]

Internet-Draft SIP Benchmarking Terminology November 2012

 Issues:

 The issue of overload in SIP networks is currently a topic of

 discussion in the SIPPING WG. The normal response to an overload

 stimulus -- sending a 503 response -- is considered inadequate and

 new response codes and behaviors may be specified in the future.

 From the perspective of this document, all these responses will be

 considered to be failures. There is thus no dependency between

 this document and the ongoing work on the treatment of overload

 failure.

3.1.6. Session Attempt

 Definition:

 A SIP request sent by the EA that has not received a final

 response.

 Discussion:

 The attempted session may be Invite Initiated or Non-invite

 Initiated. When counting the number of session attempts we

 include all INVITEs that are rejected for lack of authentication

 information. The EA needs to record the total number of session

 attempts including those attempts that are routinely rejected by a

 proxy that requires the UA to authenticate itself. The EA is

 provisioned to deliver a specific number of session attempts per

 second. But the EA must also count the actual number of session

 attempts per given tie interval. Is there a time limit to declaring a failed attempt?

 Measurement Units:

 N/A.

 Issues:

 None.

 See Also:

 Session

 Session Attempt Rate

 Invite-initiated Session

 Non-Invite initiated Session

3.1.7. Established Session

 Definition:

 A SIP session for which the EA acting as the UE/UA has received a

 200 OK message.

Davids, et al. Expires May 12, 2013 [Page 20]

Internet-Draft SIP Benchmarking Terminology November 2012

 Discussion:

 An Established Session MAY be Invite Initiated or Non-invite

 Initiated.

 Measurement Units:

 N/A.

 Issues:

 None.

 See Also:

 Invite-initiated Session

 Session Attempting State

 Session Disconnecting State

3.1.8. Invite-initiated Session (IS)

 Definition:

 A Session that is created by an exchange of messages in the

 Signaling Plane, the first of which is a SIP INVITE request.

 Discussion:

 When an IS becomes an Established Session its signaling component

 is identified by the SIP dialog parameter values, Call-ID, To-tag,

 and From-tag (RFC3261 [RFC3261]). An IS may have zero, one or

 multiple Associated Media descriptions in the SDP body. The

 inclusion of media is test case dependent. An IS is successfully

 established if the following two conditions are met:

 1. Sess.sig is established by the end of Establishment Threshold

 Time (c.f. Section 3.3.3), and

 2. If a media session is described in the SDP body of the

 signaling message, then the media session is established by

 the end of Establishment Threshold Time (c.f. Section 3.3.3).

 An SBC or B2BUA may receive media from a calling or called

 party before a signaling dialog is established and certainly

 before a confirmed dialog is established. The EA can be built

 in such a way that it does not send early media or it needs to

 include a parameter that indicates when it will send media.

 This parameter must be included in the list of test setup

 parameters in Section 5.1 of [I-D.ietf-bmwg-sip-bench-meth]

 Measurement Units:

 N/A.

Davids, et al. Expires May 12, 2013 [Page 21]

Internet-Draft SIP Benchmarking Terminology November 2012

 Issues:

 None.

 See Also:

 Session

 Non-Invite initiated Session

 Associated Media

3.1.9. Non-INVITE-initiated Session (NS)

 Definition:

 A session that is created by an exchange of SIP messages in the

 Signaling Plane the first of which is not a SIP INVITE message.

 Discussion:

 An NS is successfully established if the Session Attempt via a

 non- INVITE request results in the EA receiving a 2xx reply before

 the expiration of the Establishment Threshold timer (c.f.,

 Section 3.3.3). An example of a NS is a session created by the

 SUBSCRIBE request.

 Measurement Units:

 N/A.

 Issues:

 None.

 See Also:

 Session

 Invite-initiated Session

3.1.10. Session Attempt Failure (is this the same as Session Attempt?) there was language “routinely rejected by a

 proxy that requires the UA to authenticate itself. “

 Definition:

 A session attempt that does not result in an Established Session.

 Discussion:

 The session attempt failure may be indicated by the following

 observations at the EA:

 1. Receipt of a SIP 4xx, 5xx, or 6xx class response to a Session

 Attempt.

 2. The lack of any received SIP response to a Session Attempt

 within the Establishment Threshold Time (c.f. Section 3.3.3).

Davids, et al. Expires May 12, 2013 [Page 22]

Internet-Draft SIP Benchmarking Terminology November 2012

 Measurement Units:

 N/A.

 Issues:

 None.

 See Also:

 Session Attempt

3.1.11. Standing Sessions Count

 Definition:

 The number of Sessions currently established on the DUT/SUT at any

 instant.

 Discussion:

 The number of Standing Sessions is influenced by the Session

 Duration and the Session Attempt Rate. Benchmarks MUST be

 reported with the maximum and average Standing Sessions for the

 DUT/SUT for the duration of the test. In order to determine the

 maximum and average Standing Sessions on the DUT/SUT for the

 duration of the test it is necessary to make periodic measurements

 of the number of Standing Sessions on the DUT/SUT. The

 recommended value for the measurement period is 1 second. Since

 we cannot directly poll the DUT/SUT, we take the number of

 standing sessions on the DUT/SUT to be the number of distinct

 calls as measured by the number of distinct Call-IDs that the EA

 is processing at the time of measurement. The EA must make that

 count available for viewing ad recording.

 Measurement Units:

 Number of sessions

 Issues:

 None.

 See Also:

 Session Duration

 Session Attempt Rate

 Session Attempt Rate

 Emulated Agent

3.2. Test Components

Davids, et al. Expires May 12, 2013 [Page 23]

Internet-Draft SIP Benchmarking Terminology November 2012

3.2.1. Emulated Agent

 Definition:

 A device in the test topology that initiates/responds to SIP

 messages as one or more session endpoints and, wherever

 applicable, sources/receives Associated Media for Established

 Sessions.

 Discussion:

 The EA functions in the Signaling and Media Planes. The Tester

 may act as multiple EAs.

 Measurement Units:

 N/A

 Issues:

 None.

 See Also:

 Media Plane

 Signaling Plane

 Established Session

 Associated Media

3.2.2. Signaling Server

 Definition:

 Device in the test topology that acts to create sessions between

 EAs. This device is either a DUT or a component of a SUT.

 Discussion:

 The DUT MUST be an RFC 3261 capable network equipment such as a

 Registrar, Redirect Server, User Agent Server, Stateless Proxy, or

 Stateful Proxy. A DUT MAY also include B2BUA or SBC.

 Measurement Units:

 NA

 Issues:

 None.

 See Also:

 Signaling Plane

3.2.3. SIP-Aware Stateful Firewall

Davids, et al. Expires May 12, 2013 [Page 24]

Internet-Draft SIP Benchmarking Terminology November 2012

 Definition:

 Device in the test topology that provides protection against

 various types of security threats to which the Signaling and Media

 Planes of the EAs and Signaling Server are vulnerable.

 Discussion:

 Threats may include Denial-of-Service, theft of service and misuse

 of service.he SIP-Aware Stateful Firewall MAY be an internal

 component or function of the Session Server. The SIP-Aware

 Stateful Firewall MAY be a standalone device. If it is a

 standalone device it MUST be paired with a Signaling Server. If

 it is a standalone device it MUST be benchmarked as part of a SUT.

 SIP-Aware Stateful Firewalls MAY include Network Address

 Translation (NAT) functionality. Ideally, the inclusion of the

 SIP-Aware Stateful Firewall in the SUT does not lower the measured

 values of the performance benchmarks.

 Measurement Units:

 N/A

 Issues:

 None.

 See Also:

3.2.4. SIP Transport Protocol

 Definition:

 The protocol used for transport of the Signaling Plane messages.

 Discussion:

 Performance benchmarks may vary for the same SIP networking device

 depending upon whether TCP, UDP, TLS, SCTP, or another transport

 layer protocol is used. For this reason it MAY be necessary to

 measure the SIP Performance Benchmarks using these various

 transport protocols. Performance Benchmarks MUST report the SIP

 Transport Protocol used to obtain the benchmark results.

 Measurement Units:

 TCP,UDP, SCTP, TLS over TCP, TLS over UDP, or TLS over SCTP

 Issues:

 None.

Davids, et al. Expires May 12, 2013 [Page 25]

Internet-Draft SIP Benchmarking Terminology November 2012

 See Also:

3.3. Test Setup Parameters

3.3.1. Session Attempt Rate

 Definition:

 Configuration of the EA for the number of sessions per second that

 the EA attempts to establish using the services of the DUT/SUT.

 Discussion:

 The Session Attempt Rate is the number of sessions per second that

 the EA sends toward the DUT/SUT. Some of the sessions attempted

 may not result in a session being established. A session in this

 case may be either an IS or an NS.

 Measurement Units:

 Session attempts per second

 Issues:

 None.

 See Also:

 Session

 Session Attempt

**3.3.2. IS Media Attempt Rate**

 Definition:

 Configuration on the EA for the rate, measured in sessions per

 second, at which the EA attempts to establish INVITE-initiated

 sessions with Associated Media, using the services of the DUT/SUT.

 Discussion:

 An IS is not required to include a media description. The IS

 Media Attempt Rate defines the number of media sessions we are

 trying to create, not the number of media sessions that are

 actually created. Some attempts might not result in successful

 sessions established on the DUT.

 Measurement Units:

 session attempts per second (saps)

 Issues:

Davids, et al. Expires May 12, 2013 [Page 26]

Internet-Draft SIP Benchmarking Terminology November 2012

 None.

 See Also:

 IS

3.3.3. Establishment Threshold Time

 Definition:

 Configuration of the EA for representing the amount of time that

 an EA will wait before declaring a Session Attempt Failure.

 Discussion:

 This time duration is test dependent.

 It is RECOMMENDED that the Establishment Threshold Time value be

 set to Timer B (for ISs) or Timer F (for NSs) as specified in RFC

 3261, Table 4 [RFC3261]. Following the default value of T1

 (500ms) specified in the table and a constant multiplier of 64

 gives a value of 32 seconds for this timer (i.e., 500ms \* 64 =

 32s).

 Measurement Units:

 seconds

 Issues:

 None.

 See Also:

 session establishment failure

3.3.4. Session Duration

 Definition:

 Configuration of the EA that represents the amount of time that

 the SIP dialog is intended to exist between the two EAs associated

 with the test.

 Discussion:

 The time at which the BYE is sent will control the Session

 Duration

 Normally the Session Duration will be the same as the Media

 Session Hold Time. However, it is possible that the dialog

 established between the two EAs can support different media

 sessions at different points in time. Providing both parameters

 allows the testing agency to explore this possibility.

Davids, et al. Expires May 12, 2013 [Page 27]

Internet-Draft SIP Benchmarking Terminology November 2012

 Measurement Units:

 seconds

 Issues:

 None.

 See Also:

 Media Session Hold Time

3.3.5. Media Packet Size

 Definition:

 Configuration on the EA for a fixed size of packets used for media

 streams.

 Discussion:

 For a single benchmark test, all sessions use the same size packet

 for media streams. The size of packets can cause variation in

 performance benchmark measurements. Know that media is not being measured here, but would it be reasonable to have media fixed sizes for voice and video?

 Measurement Units:

 bytes

 Issues:

 None.

 See Also:

3.3.6. Media Offered Load

 Definition:

 Configuration of the EA for the constant rate of Associated Media

 traffic offered by the EA to the DUT/SUT for one or more

 Established Sessions of type IS.

 Discussion:

 The Media Offered Load to be used for a test MUST be reported with

 three components:

 1. per Associated Media stream;

 2. per IS;

 3. aggregate.

 For a single benchmark test, all sessions use the same Media

 Offered Load per Media Stream. There may be multiple Associated

 Media streams per IS. The aggregate is the sum of all Associated

 Media for all IS.

Davids, et al. Expires May 12, 2013 [Page 28]

Internet-Draft SIP Benchmarking Terminology November 2012

 Measurement Units:

 packets per second (pps)

 Issues:

 None.

 See Also:

 Established Session

 Invite Initiated Session

 Associated Media

3.3.7. Media Session Hold Time

 Definition:

 Parameter configured at the EA, that represents the amount of time

 that the Associated Media for an Established Session of type IS

 will last.

 Discussion:

 The Associated Media streams may be bi-directional or uni-

 directional as indicated in the test methodology.

 Normally the Media Session Hold Time will be the same as the

 Session Duration. However, it is possible that the dialog

 established between the two EAs can support different media

 sessions at different points in time. Providing both parameters

 allows the testing agency to explore this possibility.

 Measurement Units:

 seconds

 Issues:

 None.

 See Also:

 Associated Media

 Established Session

 Invite-initiated Session (IS)

3.3.8. Loop Detection Option

 Definition:

 An option that causes a Proxy to check for loops in the routing of

 a SIP request before forwarding the request.

Davids, et al. Expires May 12, 2013 [Page 29]

Internet-Draft SIP Benchmarking Terminology November 2012

 Discussion:

 This is an optional process that a SIP proxy may employ; the

 process is described under Proxy Behavior in RFC 3261 [RFC3261] in

 Section 16.3 Request Validation and that section also contains

 suggestions as to how the option could be implemented. Any

 procedure to detect loops will use processor cycles and hence

 could impact the performance of a proxy.

 Measurement Units:

 NA

 Issues:

 None.

 See Also:

3.3.9. Forking Option

 Definition:

 An option that enables a Proxy to fork requests to more than one

 destination.

 Discussion:

 This is a process that a SIP proxy may employ to find the UAS.

 The option is described under Proxy Behavior in RFC 3261 in

 Section 16.1. A proxy that uses forking must maintain state

 information and this will use processor cycles and memory. Thus

 the use of this option could impact the performance of a proxy and

 different implementations could produce different levels of impact.

 SIP supports serial or parallel forking. When performing a test,

 the type of forking mode MUST be indicated.

 Measurement Units:

 The number of endpoints that will receive the forked invitation.

 A value of 1 indicates that the request is destined to only one

 endpoint, a value of 2 indicates that the request is forked to two

 endpoints, and so on. This is an integer value ranging between 1

 and N inclusive, where N is the maximum number of endpoints to

 which the invitation is sent.

 Type of forking used, namely parallel or serial.

 Issues:

 None.

Davids, et al. Expires May 12, 2013 [Page 30]

Internet-Draft SIP Benchmarking Terminology November 2012

 See Also:

3.4. Benchmarks

3.4.1. Registration Rate

 Definition:

 The maximum number of registrations that can be successfully

 completed by the DUT/SUT in a given time period without

 registration failures in that time period.

 Discussion:

 This benchmark is obtained with zero failure in which 100% of the

 registrations attempted by the EA are successfully completed by

 the DUT/SUT. The registration rate provisioned on the Emulated

 Agent is raised and lowered as described in the algorithm in the

 companion methodology draft [I-D.ietf-bmwg-sip-bench-meth] until a

 traffic load consisting of registrations at the given attempt rate

 over the sustained period of time identified by T in the algorithm

 completes without failure.

 Measurement Units:

 registrations per second (rps)

 Issues:

 None.

 See Also:

3.4.2. Session Establishment Rate

 Definition:

 The maximum number of sessions that can be successfully completed

 by the DUT/SUT in a given time period without session

 establishment failures in that time period.

 Discussion:

 This benchmark is obtained with zero failure in which 100% of the

 sessions attempted by the Emulated Agent are successfully

 completed by the DUT/SUT. The session attempt rate provisioned on

 the EA is raised and lowered as described in the algorithm in the

 accompanying methodology document, until a traffic load at the

 given attempt rate over the sustained period of time identified by

 T in the algorithm completes without any failed session attempts.

 Sessions may be IS or NS or a mix of both and will be defined in

 the particular test.

Davids, et al. Expires May 12, 2013 [Page 31]

Internet-Draft SIP Benchmarking Terminology November 2012

 Measurement Units:

 sessions per second (sps)

 Issues:

 None.

 See Also:

 Invite-initiated Sessions

 Non-INVITE initiated Sessions

 Session Attempt Rate

3.4.3. Session Capacity

 Definition:

 The maximum value of Standing Sessions Count achieved by the DUT/

 SUT during a time period T in which the EA is sending session

 establishment messages at the Session Establishment Rate.

 Discussion:

 Sessions may be IS or NS. If they are IS they can be with or

 without media. When benchmarking Session Capacity for sessions

 with media it is required that these sessions be permanently

 established (i.e., they remain active for the duration of the

 test.) This can be achieved by causing the EA not to send a BYE

 for the duration of the testing (this is confusing, why would the EA send a BYE if test is running? I figured that a session is set-up, then media flows with that session; next sessionis set-up, then media flows, etc). In the signaling plane, this

 requirement means that the dialog lasts as long as the test lasts.

 When media is present, the Media Session Hold Time MUST be set to

 infinity so that sessions remain established for the duration of

 the test. If the DUT/SUT is dialog-stateful, then we expect its

 performance will be impacted by setting Media Session Hold Time to

 infinity, since the DUT/SUT will need to allocate resources to

 process and store the state information. The report of the

 Session Capacity must include the Session Establishment Rate at

 which it was measured.

 Measurement Units:

 sessions

 Issues:

 None.

 See Also:

 Established Session

Davids, et al. Expires May 12, 2013 [Page 32]

Internet-Draft SIP Benchmarking Terminology November 2012

 Session Attempt Rate

 Session Attempt Failure

3.4.4. Session Overload Capacity

 Definition:

 The maximum number of Established Sessions that can exist

 simultaneously on the DUT/SUT until it stops responding to Session

 Attempts.

 Discussion:

 Session Overload Capacity is measured after the Session Capacity

 is measured. The Session Overload Capacity is greater than or

 equal to the Session Capacity. When benchmarking Session Overload

 Capacity, continue to offer Session Attempts to the DUT/SUT after

 the first Session Attempt Failure occurs and measure Established

 Sessions until there is no SIP message response for the

 duration of the Establishment Threshold. Note that the Session

 Establishment Performance is expected to decrease after the first

 Session Attempt Failure occurs (again, I think system recovery is important benchmark).

 Units:

 Sessions

 Issues:

 None.

 See Also:

 Overload

 Session Capacity

 Session Attempt Failure

3.4.5. Session Establishment Performance

 Definition:

 The percent of Session Attempts that become Established Sessions

 over the duration of a benchmarking test.

 Discussion:

 Session Establishment Performance is a benchmark to indicate

 session establishment success for the duration of a test. The

 duration for measuring this benchmark is to be specified in the

 Methodology. The Session Duration SHOULD be configured to

 infinity so that sessions remain established for the entire test

 duration.

Davids, et al. Expires May 12, 2013 [Page 33]

Internet-Draft SIP Benchmarking Terminology November 2012

 Session Establishment Performance is calculated as shown in the

 following equation:

 Session Establishment = Total Established Sessions

 Performance --------------------------

 Total Session Attempts

 Session Establishment Performance may be monitored real-time

 during a benchmarking test. However, the reporting benchmark MUST

 be based on the total measurements for the test duration.

 Measurement Units:

 Percent (%)

 Issues:

 None.

 See Also:

 Established Session

 Session Attempt

3.4.6. Session Attempt Delay

 Definition:

 The average time measured at the EA for a Session Attempt to

 result in an Established Session.

 Discussion:

 Time is measured from when the EA sends the first INVITE for the

 call-ID in the case of an IS. Time is measured from when the EA

 sends the first non-INVITE message in the case of an NS. Session

 Attempt Delay MUST be measured for every established session to

 calculate the average. Session Attempt Delay MUST be measured at

 the Session Establishment Rate (max?).

 Measurement Units:

 Seconds

 Issues:

 None.

 See Also:

 Session Establishment Rate

3.4.7. IM Rate

Davids, et al. Expires May 12, 2013 [Page 34]

Internet-Draft SIP Benchmarking Terminology November 2012

 Definition:

 Maximum number of IM messages completed by the DUT/SUT.

 Discussion:

 For a UAS, the definition of success is the receipt of an IM

 request and the subsequent sending of a final response (should a more technical definition of the request / response be given? I am not IM/SIP expert, but figure there must be a message name for each).

 For a UAC, the definition of success is the sending of an IM

 request and the receipt of a final response to it. For a proxy,

 the definition of success is as follows:

 A. the number of IM requests it receives from the upstream client

 MUST be equal to the number of IM requests it sent to the

 downstream server; and

 B. the number of IM responses it receives from the downstream

 server MUST be equal to the number of IM requests sent to the

 downstream server; and

 C. the number of IM responses it sends to the upstream client

 MUST be equal to the number of IM requests it received from

 the upstream client.

 Measurement Units:

 IM messages per second

 Issues:

 None.

 See Also:

4. IANA Considerations

 This document requires no IANA considerations.

5. 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 [RFC3261], RFC 3550 [RFC3550], RFC3711

 [RFC3711] and various other drafts. This document attempts to

 formalize a set of common terminology for benchmarking SIP networks.

 Packets with unintended and/or unauthorized DSCP or IP precedence

 values may present security issues. Determining the security

 consequences of such packets is out of scope for this document.

Davids, et al. Expires May 12, 2013 [Page 35]

Internet-Draft SIP Benchmarking Terminology November 2012

6. Acknowledgments

 The authors would like to thank Keith Drage, Cullen Jennings, Daryl

 Malas, Al Morton, and Henning Schulzrinne for invaluable

 contributions to this document. Dale Worley provided an extensive

 review that lead to improvements in the documents.

7. References

7.1. 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.

 [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.

 [I-D.ietf-bmwg-sip-bench-meth]

 Davids, C., Gurbani, V., and S. Poretsky, "Methodology for

 Benchmarking SIP Networking Devices",

 draft-ietf-bmwg-sip-bench-meth-05 (work in progress),

 October 2012.

7.2. Informational References

 [RFC2285] Mandeville, R., "Benchmarking Terminology for LAN

 Switching Devices", RFC 2285, February 1998.

 [RFC1242] Bradner, S., "Benchmarking terminology for network

 interconnection devices", RFC 1242, July 1991.

 [RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V.

 Jacobson, "RTP: A Transport Protocol for Real-Time

 Applications", STD 64, RFC 3550, July 2003.

 [RFC3711] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K.

 Norrman, "The Secure Real-time Transport Protocol (SRTP)",

 RFC 3711, March 2004.

 [I-D.ietf-soc-overload-design]

 Hilt, V., Noel, E., Shen, C., and A. Abdelal, "Design

 Considerations for Session Initiation Protocol (SIP)

Davids, et al. Expires May 12, 2013 [Page 36]

Internet-Draft SIP Benchmarking Terminology November 2012

 Overload Control", draft-ietf-soc-overload-design-08 (work

 in progress), July 2011.

 [I-D.ietf-soc-overload-control]

 Gurbani, V., Hilt, V., and H. Schulzrinne, "Session

 Initiation Protocol (SIP) Overload Control",

 draft-ietf-soc-overload-control-10 (work in progress),

 October 2012.

Appendix A. White Box Benchmarking Terminology

 Session Attempt Arrival Rate

 Definition:

 The number of Session Attempts received at the DUT/SUT over a

 specified time period.

 Discussion:

 Sessions Attempts are indicated by the arrival of SIP INVITES OR

 SUBSCRIBE NOTIFY messages. Session Attempts Arrival Rate

 distribution can be any model selected by the user of this

 document. It is important when comparing benchmarks of different

 devices that same distribution model was used. Common

 distributions are expected to be Uniform and Poisson.

 Measurement Units:

 Session attempts/sec

 Issues:

 None.

 See Also:

 Session Attempt

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

Davids, et al. Expires May 12, 2013 [Page 37]

Internet-Draft SIP Benchmarking Terminology November 2012

 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

Davids, et al. Expires May 12, 2013 [Page 38]