SPECsip Infrastructure and Application Benchmarks

Yao-Min Chen
Azeem Jiva
Madhava Dass
Oracle
500 Oracle Pkwy,
Redwood City, CA 94065,
USA
yaomin.chen@oracle.com
azeem.jiva@oracle.com
madhav.dass@oracle.com

Victoria Roxas
Sam Warner
Intel
2200 Mission College
Blvd., Santa Clara,
CA 95054, USA
victoria.roxas@intel.com
samuel.r.warner@intel.com

Gary DeVal
IBM
4205 S. Miami Blvd.
Durham, NC 27709
USA
gjdeval@us.ibm.com

Wei Chen
Kenny Lee
Voxeo Labs
800 Menlo Ave Suite 210
Menlo Park, CA 94025
USA
wchen@voxeolabs.com
klee@voxeolabs.com

ABSTRACT
We describe two SIP benchmarks developed by SPEC. SPECsip_Infrastructure2011 is a SIP Proxy benchmark with realistic user behavior modeling. The second benchmark, not yet officially named, is for a JSR-289 conforming application server. We describe the design, usage and the underlying methodologies for both benchmarks.

Categories and Subject Descriptors
B.8.2 [**Performance Analysis and Design Aids**]; D.2.8 [**Metrics**] *Performance Measures;* D.4.8 [**Performance**] *Measurement, Modeling and Prediction*

Keywords
SPEC; SIP; SPECsip_Infrastructure2011; JSR289; RFC3261

1. INTRODUCTION
Marketing trends indicate that one of the key contributing factors to the expected and observed growth of VoIP and Internet multimedia industry is to unify the communication infrastructure and therefore lowering operating costs. The simplification and the reduction of costs are accelerating the acceptance and adoption in the worldwide enterprise today. The change is happening, thus validating the importance of having standard SIP benchmarks for both infrastructure and applications.

The SIP subcommittee under Open System Group (OSG) in SPEC was formed to research, plan, and initiate the creation of benchmarks in the converging area of multimedia communication over IP that use the IETF Session Initiation Protocol (SIP) [1]. The subcommittee has released one (infrastructure) benchmark and actively developing the second (application) benchmark.

SPECsip_Infrastructure2011 [2] was released in 2011 and a subsequently revised version is being worked on, towards a 2013 release. The benchmark provides for realistic set of workloads that permits fair and accurate analysis and evaluation of platform performance, at the Service Provider (ISP, NSP, and Carriers) and Enterprise level. The benchmark models a VoIP deployment supporting RFC3261 [1] and transport UDP. It is more a macro or full-system benchmark meant to capture user behavior and to be useful for capacity planning. Thus it uses "Simultaneous Number of Supported Subscribers" as its primary performance metric.

Currently also under active development is a benchmark that will consist of an application servlet to be run on JSR-289 [3] conforming application servers, a set of HTTP and SIP load generators to be scaled to distributed, multi-client testing rig, and a FABAN [4] based harness that controls over all benchmarking.

2. INFRASTRUCTURE BENCHMARK
SPECsip_Infrastructure2011 is both a specification and a released body of code that can be run and submitted for publication using SPEC's acceptance process. The benchmark focuses on a single node SIP server system under test, rather than complete network architecture such as IMS.

2.1 Design Goals
The main principle behind the design of the SPECsip_Infrastructure2011 benchmark is to emulate SIP user behavior in a realistic manner. Users are the main focus of how SIP traffic is generated. This means defining a set of assumptions for how users behave, applying these assumptions through the use of profiles, and implementing the benchmark to reproduce and emulate these profiles.

The main goal for the benchmark is to evaluate SIP server systems as a comparative benchmark in order to aid customers in analyzing core SIP operations within the infrastructure and to aid in capacity planning and provisioning of systems. Standard SIP scenarios and the traffic profile is included in the standard default workload but can be customized to fit one's own environment behavior in ISP, Telco, University, or Enterprise.

2.2 Components
The benchmark consists of several logical components that make up the benchmark. 1) *SUT* (SIP Server for an Application), which is the combination of hardware and software that provides the SIP application (in this case, VoIP) that the benchmark stresses. The SUT performs both SIP stateful proxying that forwards SIP requests and responses between users and registration that records user location provided by users via the REGISTER request. 2) *Clients* (Workload Generators and Consumers), which emulate the users sending and receiving SIP traffic. This is the load driver part of the benchmark. The load drivers use the SIPp tool [5] under the management and control of FABAN benchmark harness. 3) *Benchmark Harness*, which is responsible for coordinating the test: starting and stopping multiple client workload generators; collecting and aggregating the results from

Copyright is held by the author/owner(s).
ICPE'13, April 21–24, 2013, Prague, Czech Republic.
ACM 978-1-4503-1636-1/13/04.

the clients; determining if a run was successful and displaying the results. It is based on the FABAN framework.

2.3 Workloads and Metrics

The benchmark generates authenticated INVITE and REGISTER transactions and tracks transaction success rate and response-time histograms. The primary benchmark metric (score) is the number of supported subscribers. The secondary metrics include transaction success rate, call success rate, statistics on call mix, and INVITE and REGISTER latencies in terms of average and percentiles at different QoS levels (250ms, 500ms, and 600ms). The success of a run is dictated by whether the call mix, success rates, and QoS are met under a given subscriber population size. The benchmark has been validated on OpenSIPS [6] and Oracle Communications Converged Application Server (OCCAS) [7].

2.4 Benchmarking for the Cloud

The benchmark can be applied to test SIP servers deployed in the cloud. A SIP server, such as OCCAS, may consist of an engine tier and a data tier, where the engine tier performs transaction processing while the data tier maintains the session state. Each tier can be a cluster of managed application servers, where each server runs on its own virtual machine (VM). The VMs can be distributed on multiple blade servers for high availability and horizontal scaling. The workload drivers are used to drive SIP load to the engine tier, which communicates with data tier through inter-VM fabric. The authentication and authorization database can run on a database cluster that runs on VMs as well.

3. Application Benchmark

This benchmark targets SIP and HTTP converged applications as well as enterprise call scenarios. Besides Java EE based application servers such as Weblogic, Websphere and JBoss, the benchmark will be applicable to many JSR289-conforming application servers. For example, Voxeo provides cloud-based telephony application API which runs on JSR289 sip application servers. Developers can add sip services to web applications to bring real-time communication to the apps.

3.1 Design Goals

The benchmark is designed with the following goals. First, it should be a system-level benchmark that represents what industry does so we should avoid functional complexity if it does not add significantly to the workload. It is desired that non-SIP experts can relate to the use cases, which are from implementations and generalized, Second, the benchmark should be standard (JSR 289) compliant, which is for converged application that use HTTP and SIP session state. The benchmark should exercise multiple SIP sessions and HTTP transactions form a complete call flow. Third, the benchmark should use high level API and programming constructs like customers of an application server would. This will be fulfilled by using JPA and JSP. We made conscious decision about JPA such that database size is moderate so that the benchmark better illustrates application server performance, not limited by the database processing capacity. The database is used to store user login info, and operational info such as the number of voice mails, length and other attributes of the voice mails. The web page construction will be based on JSPs, which is fairly standard in enterprise applications.

3.2 Call Scenarios

The application benchmark will implement the call flows corresponding to the following scenarios: Scenario 1 (No Answer): There is no answer from called party so the call is forwarded to Voice Mail. Scenario 2 (Converged Applications): A user uses Web to browse the list of voice mails and play back a voice message using a SIP call. Scenario 3 (Three-way Conferencing): A call is established and then a third party is added to the existing call. Scenario 4 (Call Forwarding): A placed call receives no answer, or a busy tone, and gets forward to another number.

3.3 Benchmark Metrics

Besides the call and transaction level statistics in SPECsip_Infrastructure2011, the Application Benchmark will collect system capacity information such as the number of concurrent calls, as well as the call volume within a given period of time. Certainly, being a benchmark for a converged application including SIP and HTTP protocols, we will add transaction statistics such as throughput and latency statistics for the HTTP protocol.

4. CONCLUSIONS

SPECsip_Infrastructure2011 has been developed, and used in various situations such as sizing, benchmarking, and processor performance projection. The benchmark is used on traditional platforms as well as in cloud deployment. The actively developed Application benchmark targets enterprise applications and includes both HTTP and SIP loads, and a rich set of call scenarios. To reduce complexity and speed up development, we currently do not model Unified Messaging but may consider it in future enhancement. We also do not model RTP part of a VoIP call but may consider adding RTP modeling, based on JSR309 [8], in the future.

5. REFERENCES

[1] 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.

[2] Standard Performance Evaluation Corporation, "SPECsip_Infrastructure2011", http://www.spec.org/sipinf2011.

[3] Java Community Process, "JSR-000289 SIP Servlet 1.1", http://jcp.org/aboutJava/communityprocess/final/jsr289.

[4] "FABAN Harness and Benchmark Framework", http://java.net/projects/faban.

[5] SIPp, http://sipp.sourceforge.net.

[6] "The OpenSIPS Project", http://www.opensips.org.

[7] "Oracle Communications Converged Application Server", http://www.oracle.com.

[8] Java Community Process, "JSR-000309 Media Server Control API", http://jcp.org/aboutJava/communityprocess/final/jsr309.