White Paper

Understanding the New Performance Benchmark for IP Multimedia Subsystem (IMS) Architecture to Enable Next-Generation Networking (NGN)

Table of Contents

List of Figures

List of Tables

Executive Summary

The IP Multimedia Subsystem (IMS) framework is part of the 3rd Generation Partnership Project (3GPP) standard architecture and protocol specification for deploying real-time IP multimedia services in mobile networks.

TISPAN – the Telecoms & Internet Converged Services & Protocols for Advanced Networks group, a standardization body of the European Telecommunications Standards Institute (ETSI) – has extended IMS to support the deployment of IP services for all types of communications networks, including fixed, cable and mobile networks. This extended support enables telecom equipment manufacturers (TEMs) and service providers (SPs) to address many of the technological changes currently taking place in the telecommunications world.

Some of the more significant changes occurring today include:

- The evolution of "traditional" wireline telecom standards to Voice over IP (VoIP) standards, with Session Initiating Protocol (SIP) as the signaling protocol
- The evolution of Global System for Mobile Communications (GSM) and Code Division Multiple Access (CDMA) networks to 3GPP and 3GPP2 standards, such as Universal Mobile Telecommunications System (UMTS) technology
- Fixed-mobile convergence through the various access technologies that have been standardized by TISPAN

As customers move to deploy IMS networks, service providers and their supporting ecosystems – TEMs, computer OEMs, systems integrators and independent software vendors (ISVs) – face the dual challenge of understanding IMS workloads and engineering those workloads for deployment. Benchmark tests will prove invaluable to them for purposes of comparison, for example, comparing the performance of two products, as well as for the purpose of predicting performance; for example, the configuration specified for a benchmark test is similar enough to a service provider's requirements that the test results can be used to estimate the performance of the deployed system.

Computing benchmarks, as well as existing models used in legacy telephony networks – such as Erlang tables, 3 minute average holding time and 1 busy hour call (BHC) per subscriber – are insufficient for those purposes. SPs and the ecosystem need IP-based models that are similar to those used for data networks and application servers. Vendors and customers stand to benefit from having an industry-standard IMS benchmark.

This white paper describes the first release of the IMS benchmark developed by the ETSI TISPAN working group. It provides an in-depth explanation of the benchmark architecture, discusses many of the core concepts, and presents a set of sample test results for illustration purposes.

NGN/IMS Overview

The following diagram (Figure 1: NGN/IMS TISPAN Architecture) depicts the IMS reference architecture. The various architectural components are the primary building blocks, which are either defined by the IMS standard, or defined by external standards and referenced by IMS. The links between the primary building blocks represent reference points over which the building blocks communicate with each other.

The IMS reference architecture is a logical architecture; it does not map functional elements to hardware or software components. Conversely, IMS products deployed in the real world do not factor neatly into the elements of the reference architecture. This fact complicates the process of comparing similar products using a benchmark.

Proceeding from a subsystem description to a benchmark test requires the presence of a complete description of all aspects of the subsystem relevant to the benchmark's performance. This description is called the system configuration, or the system under test (SUT) configuration. The description enumerates the elements of the reference architecture and enumerates all reference points that are external to the subsystem. (Reference points between elements within the subsystem are "internal.") Version 1 of the benchmark specification focuses on the Session Control Subsystem (SCS), which is made up of the Call Session Control Function (CSCF) and the User Profile Server Function (UPSF) as shown in Figure 1.

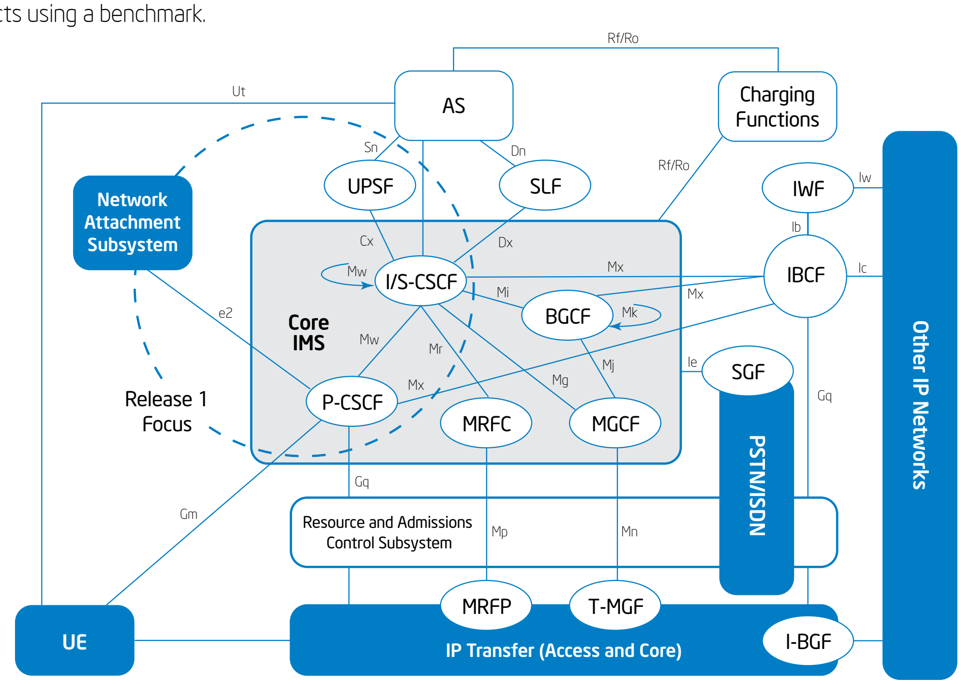

Figure 1 NGN/IMS TISPAN Architecture

The CSCF establishes, monitors, supports and releases multimedia sessions, and it manages user service interactions. The CSCF can act as a Proxy CSCF (P-CSCF), as a Serving CSCF (S-CSCF) or as an Interrogating CSCF (I-CSCF). The P-CSCF is the first point of contact for the user equipment (UE), also called the user-endpoint, within the IMS network; the S-CSCF handles the actual session states in the network; and the I-CSCF is the main point of contact within an operator's network for all IMS connections destined for a subscriber of that network operator or destined for a roaming subscriber located within that network operator's service area.

The UPSF is similar to the Home Subscriber Server (HSS) in 3GPP in that it is not part of the "core IMS." However, it exchanges information with the CSCF for functions such as routing information retrieval, authorization, authentication and filter control.

Overview of IMS Benchmark

The ETSI TS 186.008 is a technical specification composed of three parts[1]:

- An overall benchmark description, which includes environment, architectures, processes and information models that are common to all specific benchmarking scenarios
- The IMS and ETSI TISPAN SUT configurations, use-cases and scenarios, along with scenario-specific metrics and design objectives and SUT configuration parameters
- A defined initial benchmark test that specifies a traffic set, traffic profile and benchmark test procedures

As mentioned earlier in this document, Release 1 of the benchmark specification focuses on the Session Control Subsytem, or SCS; it consists of the three main CSCF elements (Proxy, Serving and Interrogating) and the UPSF, much like the HSS in 3GPP. The IMS elements that are not part of this focus are regarded as part of the test environment. Additional subsystems may be covered by future versions. Depending on the objective of the benchmark, the SUT being considered may not be the whole SCS, but rather the subsystem implementing only one of the UPSF or CSCF elements.

In Release 1 of the IMS benchmark, the following three IMS events are considered for benchmarking:

- Registration and de-registration, covering nine scenarios
- Session set-up or tear-down, covering 25 scenarios
- Page-mode messaging, covering two scenarios

Benchmark Architecture

The following diagram (Figure 2: High-Level Architecture) provides a high-level view of the IMS benchmark architecture, which consists of a test system and the system under test, or SUT. The test system emulates the user-endpoints (UEs), which issue IMS events (such as registration and de-registration, session set-up or tear-down and messaging) to the SUT.

The SUT in turn responds to these events. The test system maintains a transaction state table for each UE. Each time the test system receives a response from the SUT, it identifies that response with a UE, validates the response, updates the transaction state table and, if necessary, processes a response.

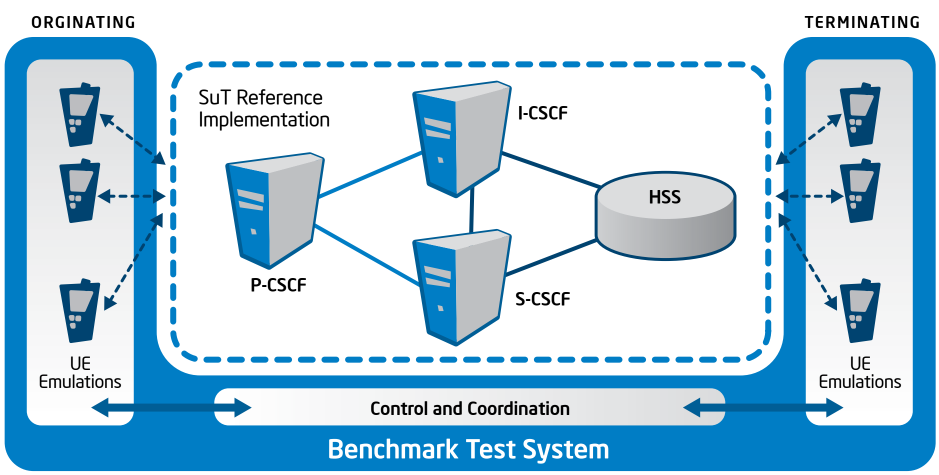

Figure 2 High-Level Architecture

[1] The "IMS/NGN Performance Benchmark" specification (ETSI TS 186.008) can be downloaded from the ETSI website at: http://pda.etsi.org/pda/queryform.asp (search for 186008 keyword):
Part 1: Core Concepts: ts_18600801v010101p.pdf
Part 2: Subsystem Configurations and Benchmarks: ts_18600802v010101p.pdf
Part 3: Traffic Sets and Traffic Profiles: ts_18600803v010101p.pdf

Defining User-Endpoints/Users

A user is a state machine running a scenario. A user may:

- Be a "callee" or a "caller"
- Create one or more calls
- Be reused to create other calls
- Randomly call any other user

A user has "use-cases" in mind that consist of a collection of scenarios, each of which describes a possible interaction determined by the behavior of the user and the system under test.

Understanding Scenarios

A scenario is a portion of an IMS event such as registration, de-registration or text messaging. A scenario is a trace of a path through a use-case. It is analogous to "call attempt" but applies to all interactions within an IMS network, such as registrations, text messages and application interactions.

A scenario can have one of three results; it can succeed, it can fail, or it can succeed functionally but take longer than allowed by the time thresholds associated with its use-case. In the latter two instances, it is deemed an "inadequately handled scenario attempt" (IHSA).

A collection of scenarios define a traffic set. Some examples of traffic sets are depicted in the table that follows (Table 1 Table 1: Traffic Set Examples).

Figure 3: Benchmark Information Model) illustrates the concepts behind the use cases, the traffic sets and the benchmark tests.

Test Scenario	Scenario ID	Type	Scenario % of System Load	Scenario Arrival Distribution	Scenario Duration Distribution (calls), message size (text messaging)
SCENARIO 9 Abandoned Call Resource reservation on both sides	PX_S2_9	float	3%	Poisson, mean selected by traffic profile	Exponential, mean 15 sec
SCENARIO 10 Abandoned Call – No resource reservation on originating side	PX_S2_10	float	3%	Poisson, mean selected by traffic profile	Exponential, mean 15 sec
SCENARIO 11 Abandoned Call – No resource reservation on terminating side	PX_S2_11	float	3%	Poisson, mean selected by traffic profile	Exponential, mean 15 sec
SCENARIO 12 Abandoned Call – No resource reservation on either side	PX_S2_12	float	3%	Poisson, mean selected by traffic profile	Exponential, mean 15 sec

Table 1: Traffic Set Examples

This benchmark standard uses the terms "scenario attempt" and "scenario attempts per second" (SAPS) rather than "call attempt" and "call attempts per second" because IMS is a transaction-oriented system that encompasses transactions of a variety of types (for example, calls, registrations, de-registrations and text messages). The more generalized term is necessary because traffic sets, as well as the real world, don't operate according to only one transaction type. Attempting to report the capacity of a system in "call attempts per second" or "registration attempts per second" for system loads that are other than purely call attempts, registration attempts and so forth, would be incorrect and misleading.

Scenario attempts could be further categorized into call-dependent (for example, conversational services or streaming services) and call-independent (for example, registration or roaming) scenario attempts. This categorization is meaningful only for network elements that can differentiate both scenario categories (for example, P-CSCF).

Each scenario is documented by the associated message flow, design objectives and metrics or measurements to be collected if that scenario is selected. Typical metrics include scenario outcome, response times, message rates and the number of inadequately handled scenarios (IHS). If these exceed a certain frequency, it is interpreted as a probability of inadequately handled scenario attempts. The SUT reaches its Design Objective Capacity (DOC) when the IHSAs exceed the design objective.

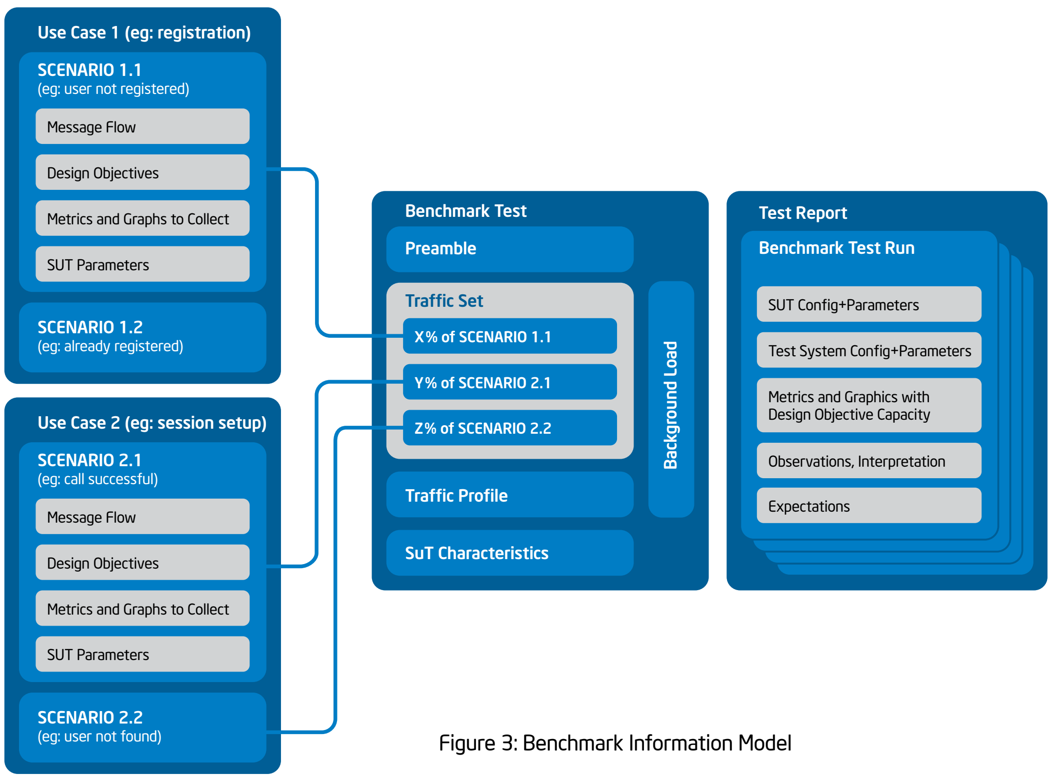

Figure 3: Benchmark Information Model

The goal of the IMS benchmark is to express a system's performance using a single "figure of merit," as is done in the legacy telephone model. To accomplish this, the "load unit" is the "scenario attempt per second" (abbreviated as SAPS) metric, applicable to any scenario in any use-case.

The heart of the benchmark test is the traffic set, which is a collection of scenarios determined to be likely to occur in a real-world situation. Within a traffic set, each scenario has an associated relative occurrence frequency, interpreted as its probability of occurring in the course of the test procedure.

Selected scenarios are those with a relative frequency higher than 0 %; these typically derive from all three use-cases.

The IMS benchmark test is also characterized by an "arrival distribution," which describes the arrival rate of occurrences of scenarios from the traffic set; and a "traffic profile," which describes the evolution of the average arrival rate as a function of time over the duration of the test procedure. The following table (Table 2) shows an example of an initial benchmark traffic-time profile.

Traffic Profile Parameter	Traffic Profile Value	Notes
PX_SimultaneousScenarios (SIMS)	2	Maximum per UE
PX_TotalProvisionedSubsribers	100,000 Subs	Data in part 2
PX_PercentRegisteredSubscribers	40%	At test start. The percentage of registered subscibers will fluctuate during the test.
PX_PercentRoamingSubscribers	None	No roaming in release 1
PX_StepNumber	3 Steps	DOC underload, DOC, and DOC overload
PX_StepTransientTime	120 Seconds	Maximum
PX_StepTime	30 Minutes	Minimum
PX_BackgroundLoad	None	
PX_SApSIncreaseAmount	10 SApS	Maximum Report three results, step before, DOC and step after
PX_SystemLoad	DOC	Reported result in scenario attempts per second
PX_IHS % InAdequatedly Handle Scenario Attempts Maximum (IHS)	0.1%	Average over a test step

Table 2: Initial Benchmark Traffic-time Profile

A "test report" is a document that, along with accompanying data files, fully describes the execution of a benchmark test on a test system. The SUT and test system, as well as their parameters, are described in sufficient detail that an independent test site can replicate the test. The test results include charts and data sets depicting the behavior of the SUT over the duration of the test.

A typical test sequence starts with an underloaded system, which is brought to its Design Objective Capacity, or DOC, and maintained at that load for a certain time. The time during which a system runs at its DOC must be long enough to provide meaningful data and highlight possible performance issues, such as memory leaks or overloaded message queues. An important indicator is the proportion of IHSAs (scenario attempts that either fail or succeed only after a time threshold) during the various phases of the test. In this example, a performance requirement is that the portion of IHSAs doesn't exceed 0.1%, averaged out over the time while the system is running at its DOC.

Test System

The test system is used to generate the appropriate load on the SUT. The benchmark specification does not mandate the use of a specific test system; however, the details of the test system must be specified in the benchmark report.

The following diagram (Figure 4: Test System and SUT Interactions) depicts the test system connections and interactions with an SUT.

The test system serves three main functions: traffic generation, network emulation and synchronization.

- Traffic generation: The test system must be able to execute use-case scenarios in accordance with the traffic-time profile. It must also be able to reproduce the appropriate traffic set, namely, a mix of scenarios with a weight for each scenario.
- Network emulation: Optional network characteristics on the various interfaces must be emulated by the test system. This includes network bandwidth, latency and error rate. These characteristics are to be set separately for each direction so that non-symmetric interfaces can be emulated (for example, up and down bandwidth on a DSL link).
- Synchronization: In instances where protocol information elements must be passed between SUT interfaces and the test system is different for those interfaces, a synchronization mechanism must exist to pass those information elements between the test systems.

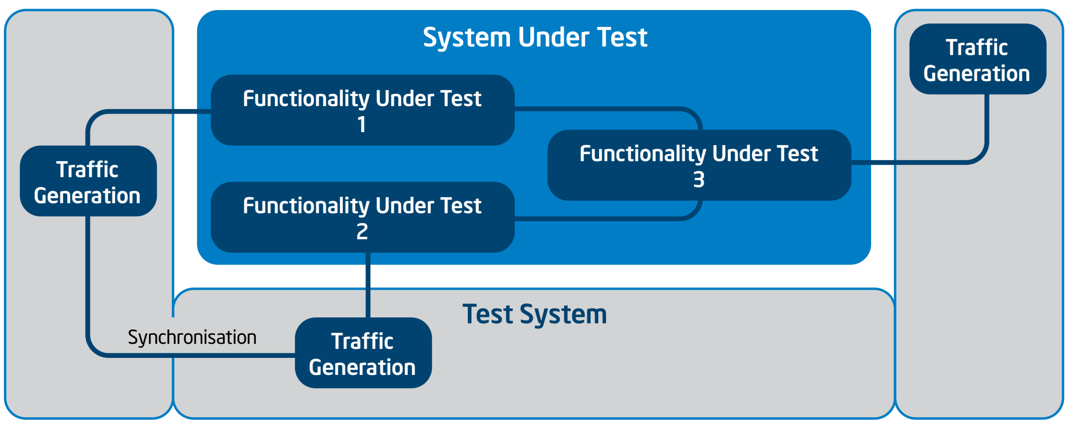

Figure 4 Test System and SUT Interactions

System under Test

An IMS/NGN benchmark must enable not only the benchmarking of a complete IMS network (as depicted in Figure 1: NGN/IMS TISPAN Architecture), but also the benchmarking of network subsystems corresponding to discrete products that may be available from a supplier. To address this requirement, the IMS benchmark standard defines a series of subsystems that can serve as an SUT for a benchmark test. IMS/NGN elements that do not appear in a subsystem are regarded as part of the test environment; these elements must be present for a subsystem to function, but the overall test environment is not itself subject to benchmarking. The following table outlines the network subsystems for which benchmarks are to be specified.

IMS/NGN Performance Benchmark Subsystem	Included 3GPP IMS Functionality	Included TISPAN NGN Functionality	Test Environment Functionality
Session Control Subsystem (SCS)	P-CSCF, I/S-CSCF, HSS	P-CSCF, I/S-CSCF, S-CSCF, UPSF	DNS, access network (e.g. SPDF, C-BGF, A-RACF, DSLAM, SBC, switches, routers)
HSS/UPSF Subsystem	HSS	UPSF	
P-CSCF Subsystem	P-CSCF	P-CSCF	DNS, access network (e.g. SPDF, C-BGF, A-RACF, DSLAM, SBC, switches, routers)
I/S-CSCF Subsystem	I-CSCF, S-CSCF	I-S/CSCF	DNS, access network (e.g. SPDF, C-BGF, A-RACF, DSLAM, SBC, switches, routers)

NOTE: The last column of Table 1 represents the elements of the test environment. In Release 1, only benchmark configurations with one network are specified; in such a configuration, DNS queries are cached locally, and hence have no significant effect on the measured metrics. Similarly in Release 1, IPv6, network errors, and network delays are not specified in benchmarks, and hence have no impact.

Table 3: Subsystems for which Benchmarks are to be Specified

For the purposes of benchmarking, however, certain rules concerning subsystem configurations are required. These rules help ensure that benchmark measurements taken from equivalent subsystems of various vendors are comparable with one another.

The general guidelines for defining an SUT configuration are as follows:

- All functional elements of the subsystem must be present in the SUT configuration
- All hardware elements used in the implementation of the SUT configuration must be completely enumerated
- All quality of service (QoS) spec measurements defined at the interfaces to the SUT must be collected as specified in the benchmark test
- All hardware-specific measurements (for example, CPU utilization, memory utilization and fabric bandwidth) specified in the benchmark test must be collected for all hardware elements used in the implementation of the SUT configuration
- SUT interface characteristics must be specified so that they can be emulated by the test system, including:
- Security, for example, IP security (IPSec), Transport Layer Security (TLS) and Datagram TLS (DTLS)
- Interface network characteristics, for example, up and down bandwidth and up and down latency

Test Procedure

A benchmark test defines the following four elements:

- A "preamble" period, which is the sequence of actions required to initialize a test system and SUT to perform a benchmark
- A "traffic set," which is the set of scenarios that simulated users perform during the test procedure, together with the relative frequency with which the scenarios occur during the test procedure
- An "arrival distribution," which describes the arrival rate of occurrences of scenarios from the traffic set
- The "traffic-time profile," which describes the evolution of the average arrival rate as a function of time over the duration of the test procedure

During the test procedure, scenarios are selected for execution. The time between the execution of subsequent scenarios is determined by the arrival distribution, and the arrival distribution is parameterized by the average arrival rate. The scenario arrival rate changes over time according to the traffic-time profile during the test procedure.

The test procedure is carried out as follows:

- The test system performs the preamble, during which any actions required to initialize the test system and the SUT are carried out. These actions generally include loading a subscriber base with subscriber data, performing transactions on the subscriber base to randomize the data, and causing the

SUT to have "packets in flight" in its internal queues, to make its state approximate the case in which it ran in a real-world deployment for some extended amount of time.

- The test system sets the initial average arrival rate to the initial value specified by the traffic-time profile. The test system delays for a random interval (calculated by the arrival distribution to achieve the average arrival rate), then randomly selects a scenario from the traffic set with a probability equal to the scenario percent of system load. This scenario then starts to run.
- As time elapses during the test procedure, the profile will change by the SAPS increase amount. When the value changes, the inter-arrival time of scenario selection (and hence system load) will change.
- When the entire traffic-time profile has been performed and the total time of the test procedure has elapsed, the test system stops sending scenarios for execution. When the test system completes executing all scenarios, the test procedure terminates.

Benchmark Test Results

The performance of Intel® Architecture-based systems running IMS workloads from generation to generation is presented in the following table (Table : Performance of Subsequent Generations of Intel Platforms). These results have been collected using the Intel IMS Bench SIPp tool acting as the test system. This tool is available online at http://sipp.sourceforge.net/ims_bench/, along with sample benchmark reports.

The traffic set used to collect these results was as follows:

- 73 percent Scenario PX_S2_4, clause 5.2.2.4: Successful Call - No resource reservation on either side
- 27 percent Scenario PX_S3_1, clause 5.3.2.1: Successful Message Exchange

A total of 100,000 subscribers were provisioned, out of which 70 percent were registered during the preamble.

SUT	DOC
Platform A	80 SAPS
Platform B	150 SAPS
Platform C	270 SAPS
Platform D	360 SAPS

Table 4: Performance of Subsequent Generations of Intel Platforms

Conclusion

This document introduced many of the concepts behind the first version of the ETSI TISPAN IMS performance benchmark.

As a brief summary, the benchmark consists of a test system that presents the system under test with workloads. The workloads consist of traffic generated by a large number of individual simulated user- endpoints, each performing an individual scenario. The collections of scenarios selected for a benchmark test define a traffic set.

The rate at which scenarios from the traffic set are attempted in the benchmark test is governed by the traffic-time profile defined for the benchmark. During the test, "inadequately handled scenario attempts" are collected and measured. When these IHSAs exceed the design objective, the system being tested has reached its Design Objective Capacity.

With the release of this specification, service providers and equipment manufacturers now have an industry-standard benchmark that can be used in two ways:

- As a predictive benchmark indicator for IMS solution performance improvement. The benchmark can be used for first-level network planning and engineering based on new processor and platform introductions being driven by Moore's Law.
- As a comparative benchmark for hardware and software IMS solution architecture selection. The benchmark provides a rule of thumb for the level of performance that should be attainable.

With the release of the IMS bench SIPp tool, an open source implementation is now available that can be used to execute the benchmark. More information on this tool can be found at http://sipp.sourceforge.net/ims_bench/.

www.intel.com

Printed in the USA 0308/MS/SD/pdf 319469-001US