 <p><b>SERVICES AND SUPPORT PROFILE</b> <b>SERVICE AVAILABILITY</b></p>	<p><b>Version 1.0: July 30, 2007</b></p>
--	--

## **Services and Support Profile –** **Service Availability**

**Version 1.0 July 30, 2007**

Copyright © 2007 SCOPE Alliance. All rights reserved.

The material contained herein is not a license, either expressed or implied, to any IPR owned or controlled by any of the authors or developers of this material or the SCOPE Alliance. The material contained herein is provided on an “AS IS” basis and to the maximum extent permitted by applicable law, this material is provided AS IS AND WITH ALL FAULTS, and the authors and developers of this material and SCOPE Alliance and its members hereby disclaim all warranties and conditions, either expressed, implied or statutory, including, but not limited to, any (if any) implied warranties that the use of the information herein will not infringe any rights or any implied warranties of merchantability or fitness for a particular purpose.

Also, there is no warranty or condition of title, quiet enjoyment, quiet possession, correspondence to description or non-infringement with regard to this material. In no event will any author or developer of this material or SCOPE Alliance be liable to any other party for the cost of procuring substitute goods or services, lost profits, loss of use, loss of data, or any incidental, consequential, direct, indirect, or special damages whether under contract, tort, warranty, or otherwise, arising in any way out of this or any other agreement relating to this material, whether or not such party had advance notice of the possibility of such damages.

Questions pertaining to this document, or the terms or conditions of its provision, should be addressed to:

SCOPE Alliance,  
c/o IEEE-ISTO  
445 Hoes Lane  
Piscataway, NJ 08854  
Attn: Board Chairman

Or

For questions or feedback, use the web-based forms found under the Contact tab on [www.scope-alliance.org](http://www.scope-alliance.org)



Version 1.0: July 30, 2007

## SERVICES AND SUPPORT PROFILE SERVICE AVAILABILITY

### 1. PURPOSE

As NEPs endeavor to engineer telecommunication systems by drawing upon the COTS ecosystem, they must have a way to qualify the components that they leverage from this environment. This gives rise to a plethora of terminologies and methodologies to perform this qualification; however, there is little consistency among the NEPs as to how they do this even though, in the end, they are attempting to accomplish the same result. Such a lack of consistency fundamentally complicates the job of ecosystem suppliers by causing them have to respond differently to each NEP. In the end, this fragmentation translates to increased costs, slower cycle times, and general confusion across this aspect of the ecosystem.

The purpose of the 'Services and Support Profile' papers is to suggest some standard terminology and methods that NEPs may converge around for portions of the carrier grade engineering qualification process.

This paper deals with the term 'Service Availability'.

### 2. AUDIENCE

The target audiences of this document are:

- ✓ Network Equipment Providers who deliver applications and solutions to telecommunications and networking carrier companies in the context of Carrier Grade Platforms
- ✓ Board, module and software vendors who market their products for use in Network Elements and other Network Equipment Provider applications built on Carrier Grade Base Platforms

### 3. REFERENCES

1. High Availability Glossary, [Unpublished] Draft 0.5, Service Availability Forum – Technical Working Group, 15-Dec-2004.
2. Reliability and Quality Measurements for Telecommunications Systems (RQMS-Wireless), GR-1929-Core Issue 2, Telcordia Technologies, February 2005.
3. Scoping the Scope: Closing the Gaps of an Open Carrier Grade Base Platform, Revision 1.1, The SCOPE Alliance (available at [www.scope-alliance.org](http://www.scope-alliance.org)) March 2007.
4. SR-332 Reliability Prediction Procedure for Electrical Equipment, Issue 1.0, Telcordia Technologies, May 2001.
5. TL9000 Quality Management System Measurements Handbook, Release 3.5, Quality Excellence for Suppliers of Telecommunications Forum – QuEST Forum, March 31, 2003.



**SERVICES AND SUPPORT PROFILE  
SERVICE AVAILABILITY**

## 4. TERMS AND DEFINITIONS

<b>CG</b>	Carrier Grade
<b>CO</b>	Central Office
<b>ETSI</b>	European Telecommunications Standards Institute
<b>FRU</b>	Field Replaceable Unit
<b>NEP</b>	Network Equipment Provider
<b>OS</b>	Operating System
<b>QoS</b>	Quality of Service
<b>SA</b>	Service Availability
<b>SAF</b>	Service Availability Forum
<b>SLA</b>	Service Level Agreement

## 5. SERVICE AVAILABILITY

Service availability is a fundamental attribute of carrier grade base platforms [3] and is a complex aspect of a system to quantify, integrate and deterministically deliver. Since service availability is a defining characteristic of a complete system, it is driven in large part by the reliability of the system's constituent components (i.e., the components integrated to construct the system) and the effectiveness of the system's fault management.

### 5.1 Definitions

The SCOPE Alliance gratefully acknowledges that, for this publication, we draw heavily upon some unpublished work of the Service Availability Forum [1].

The overarching concept of interest is *service availability* which we define as the probability that the system will properly respond to a *service* request from a user or another system with an acceptable level of *quality-of-service* as measured by the end user or client system. This includes the availability of all hardware and software elements including computation, storage and network components that comprise the *system* and which are needed in order for the service to be delivered acceptably. Service availability is the product of the availability of all elements, some which may be themselves highly available, involved in providing the service.

In order for this definition to be clear, there are some included terms that we define now.

- *Availability* in general refers to the probability that a system is in an operating condition that allows it to deliver the service(s) for which it is intended to provide. Availability is a measure that conventionally is expressed as a percentage which is calculated as the uptime divided by the uptime plus the downtime of the system. For this consideration it makes



Version 1.0: July 30, 2007

**SERVICES AND SUPPORT PROFILE  
SERVICE AVAILABILITY**

no difference whether the downtime is planned or unplanned<sup>1</sup>. Often availability is expressed as the number of nines (e.g., 99.999% is expressed as five 9's and this translates to approximately 5.26 minutes of downtime per year).

- A *service* is a set of actions that satisfy a request from a user or other system. For example, database servers satisfy requests for updates, inserts, or queries. Services may embed other subsidiary services. For example, a database-driven Web service will invoke database servers.
- *Quality of Service (QoS)* here is used to refer to all of the traits of the service that an end-user perceives as important to the acceptability of the service. In addition to the simple presence/absence of the service, this also includes aspects such as response time, accuracy, throughput, audio/video clarity, capacity, dropped calls, lost transactions, etc. QoS parameters are often defined in terms of a service level agreement (SLA) that specifies the range of acceptability.
- A *system* is an imprecise term that is often used to refer to a collection of entities. For purposes here, a system is the collection of elements, some of which may be redundant, that provide the complete context for delivering a service. Systems may embed other sub-systems. For example, a Web server may comprise a front end service group as a sub-system to manage client interactions and a back end service group as a database subsystem. Service availability, as visible to the end user, will be a function of the availability of the various embedded sub-systems used to implement the service.

With this basis, we can now extrapolate the term *carrier grade availability* which is essentially a level of *service availability* that meets telecommunications/networking carrier standards. Typically for carriers that provide mission critical services (such as voice telephony), the standards are at least a level of service availability that is 99.999% or better.

## 5.2 Allocation of Availability Budgets

As was mentioned above, it is fundamental that the expected availability of components comprising a service providing system be known and allocated in the overall process of engineering carrier grade service availability. Typically there is a relationship of dependencies in the system components (hardware, system software, middleware, application, etc.) with the overall availability a function of the combined

---

<sup>1</sup> 'Planned' is the term used to refer to downtime that can be anticipated or scheduled (such as hardware maintenance or software upgrades) while unplanned downtime is the result of system failure. In general, planned downtime is less traumatic in the sense that it can be incurred at times of reduced service implications; however, NEPs are generally not allowed to assume that planned downtime is allowed (or preferable) in the overall availability budget.



Version 1.0: July 30, 2007

**SERVICES AND SUPPORT PROFILE  
SERVICE AVAILABILITY**

availability of the components. For example, if the hardware, system software, middleware and application are each 5 9's (99.999%) available, the combination of the 4 yields a system that is only 99.996% available overall (here this is a simplification in that redundancy relationships and other factors significantly complicate an actual analysis).

To address this, a key part of the process of engineering a platform with carrier grade service availability is to establish an *availability budget*. Once again, drawing up the SAF work in [1] we define an *availability budget* to be an allocation of the allowed downtime of a system across its constituent components. The typical areas for allocation of downtime include:

- Environment: This is downtime that is triggered by factors around the system such as power, HVAC, and natural disasters.
- Operator Error: This element of downtime is caused by an operator error or mishandled procedure. Generally the size of the allocation to this area is a reflection of the complexity of the system overall as well as user interfaces such the MMI (Man Machine Interface).
- Maintenance: Most systems require some form of regular maintenance and in some cases this cannot be performed without impairing the system. A simple example of this is the installation of a new software release. Generally, NEPs engineer systems such that this component of downtime is zero (including provisions for in-service software upgrade) but that is not always the case.
- Hardware: The hardware components of the system may fail (e.g. fatigue). This includes not only the digital hardware but also related equipment such as fans and power supplies. There are sophisticated mechanisms to predict hardware reliability and often hardware failures are mitigated through redundancy (duplication of the hardware via some scheme).
- System Software: This is generally the operating system and associated software environment. It often contains the fundamental protocol stacks associated with the application in question. Generally, failures in software are mitigated by leveraging clustering environments where the software image is self-contained and isolated from other comparable service providing images (e.g., via separate hardware or virtualization).
- Platform Software: This is software that runs above the system software and provides higher level services. In the SCOPE Alliance reference architecture [1] this includes functions that orchestrate the highly available environment such as clustering, availability management, checkpointing and such.



**SERVICES AND SUPPORT PROFILE  
SERVICE AVAILABILITY**

- Application: This is actual payload software that provides the specific service for which the platform is intended to perform.

In addition, allocations commonly distinguish between service provider attributed downtime and equipment supplier (i.e. NEP) attributable downtime. For example, some environmental outages may be service provider attributable (if the service provider does not properly maintain the operational environmental conditions). Also, some operator error outages may be service provider attributable (if service provider staff does not correctly follow specified operations tasks) or may be equipment supplier attributable (if the operator interface is obscure or specified tasks are incorrect). Therefore, for purposes of this activity, the categories of operator error and environment may be (partially) irrelevant to ecosystem suppliers.

For example, if a system is expected to operate at 5 9's (approximately 300 seconds or less of downtime/year), an allocation of those 300 seconds might be as follows:

Category	Downtime Allocation	Percentage
Environment	30 seconds/year	10%
Operator Error	84 seconds/year	28%
Maintenance	30 seconds/year	10%
Hardware	36 seconds/year	12%
System Software	15 seconds/year	5%
Platform Software	30 seconds/year	10%
Application Software	75 seconds/year	25%
Total	300 seconds/year	100%

Obviously the above is a statistical average across a number of systems and a number over several years.

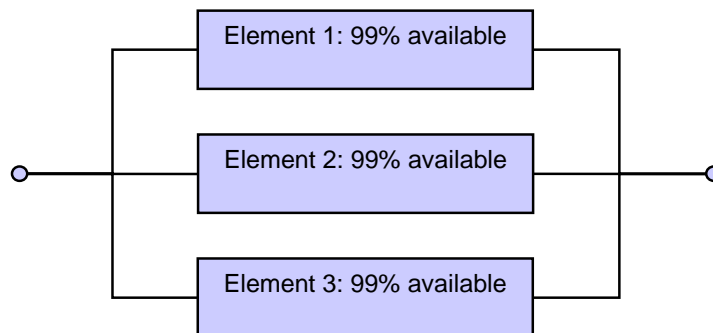
The intent of the example allocations is to standardize the categories and terminologies against which availability is budgeted and to suggest typical allocation ranges that the NEPs can agree to nominally operate within when dealing with the supplier ecosystem.

The provider ecosystem product content impacts nearly all of the above categories in some way but is most impactful on in the areas of Hardware, System Software and Platform Software since those are the typical components that NEP attempt to consume directly. It is difficult to standardize these budgets precisely for a few reasons:




**SERVICES AND SUPPORT PROFILE  
SERVICE AVAILABILITY**

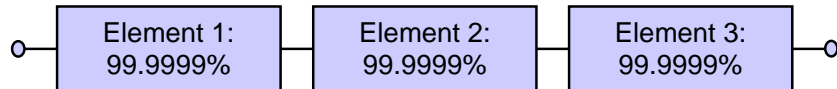
- The allocation and impact is also a function of how the NEP consumes the content.
  - A system can be composed of components that are consumed independently (e.g., one board from supplier A, another from supplier B, a chassis from supplier C, and the operating system from supplier D). A system may be consumed with some level of integration (e.g., a hardware platform with an integrated OS and HA Middleware). In this case, portions of the respective budgets above would be allocated to individual suppliers of the components.
  - When subsystems are consumed with additional integration, the supplier of the integrated platform has a much larger impact on the areas of operator error and maintenance. For this case, the supplier might be allocated the entire budget for hardware, system software, and platform software as well as some of the budget for maintenance and operator error.
- The budget allocation ratios vary depending upon the level of service availability that one is attempting to achieve (e.g., 4 9's vs. 5 9's vs. 6 9's). Because the hardware is the most predictable and consistent across this, for a lower availability system, a higher percentage of downtime might be allocated to application software and/or maintenance.
- The architecture of the system or platform that the components are being integrated into obviously plays a major role in determining the budgetary allocations. For example the following diagram depicts redundancy across three like elements and shows that a system can be made to operate with considerably higher availability than its individual components.



The resulting system is 99.9999% available if only one of the three elements must be operational at any time to provide the required level of service while it is only 99.99% available if two of the three blades must be operational.

 <p><b>SERVICES AND SUPPORT PROFILE</b> <b>SERVICE AVAILABILITY</b></p>	<p><b>Version 1.0: July 30, 2007</b></p>
--	--

Similarly, the following diagram very simply shows the effects of serial or additive relationships among elements.




All elements must be operational for the system to deliver service. Even though each element is itself 6 nines, the overall system availability is 99.9997% which is less than the individual components.

- The environmental component is very much a function of the surrounding conditions that are imposed on the system. The operator error portion is not only a function of system complexity but also staff expertise, workload, etc.

### 5.3 Methodologies and Standards

This section suggests some standard methods that can be used to quantify some of the critical parameters of service availability as overviewed in the preceding section.

1. **Hardware Reliability:** SR-332 Telcordia Technologies [4] is an accepted and relatively common method for calculating and predicting the reliability of hardware. From discussions with ecosystem suppliers, this is believed to be a conservative method for predicting the reliability of hardware components so it may be appropriate for suppliers to suggest a specific interpretation methodology.
2. **System Measurements:** The QuEST Forum of TL-9000 [5], provides quality measurement methods that can be extrapolated to characterize aspects of the service availability engineering process. In particular, Chapter 6 quantifies the definitions and measurement methods around an outage whether it be system fault induced or operator/procedural. (Note that GR-1929-Core [2] from Telcordia Technologies also characterizes outage accounting rules; however, TL-9000 appears to be a more universally accepted standard worldwide).
3. **Hardware Quality:** TL-9000 [5] also provides a methodology to assess the fielded quality of Field Replaceable Units (FRUs) across the lifecycle of product deployment (Chapter 7). This is important to standardize in order to allow objective comparisons of the “track-record” performance of suppliers as well as to set expectations for the service availability of a system as it progresses from its inception, initial deployment through to mature general deployment.
4. **Software Deployment:** TL-9000 [5] (Chapter 8) provides methodology to assess quality of software releases. One consideration here would be to allocate this set of measurements across the categories of software content summarized in the preceding section. The importance of this to software is similar to that of the pre-

 <p><b>SERVICES AND SUPPORT PROFILE</b> <b>SERVICE AVAILABILITY</b></p>	<p><b>Version 1.0: July 30, 2007</b></p>
--	--

ceding point on hardware; namely in assessing the “track-record” and life-cycle maturation of the software.

Many of these standards are encompassing and/or offer multiple variations on methods for calculations. A possible next step would be for the ecosystem suppliers to organize and suggest specific interpretation methods and/or applicable subsets.

#### **5.4 Conclusion**

This Services and Support Profile from the SCOPE Alliance has clarified the common terminology and practices that NEPs employ relative to Service Availability. In particular, it has specifically defined requisite terminology, suggested a framework for budgeting downtime, and identified the key standards/methodologies that are applicable to most NEP platform engineering activities.

A possible future work item for this area is to standardize the responsibilities for reporting events that affect service availability. Of particular interest is the responsibility of service providers to report outage events to equipment providers and for equipment providers to maintain/decompose reliability and availability statistics. The equipment providers then, in turn, will need to use this information with the component suppliers to manage performance relative to the overall service availability budget. This is necessary feedback to component providers but is often difficult to orchestrate due to confidentiality issues that NEPs must abide by with their service provider customers.