



Open Specifications for Service Availability

The Service Availability Forum and Open Specification Solutions

An Industry in Transition

Today the dependability of the communications infrastructure and the applications flowing over these networks is more important than ever. As new technologies enable powerful new services, users quickly become dependent on those services to conduct their personal and professional lives. The communications and enterprise computing industries are under enormous pressure to revamp networks and applications to accommodate explosive growth and emerging technologies and – at the same time – ensure that the services delivered are available and dependable.

New technologies and services, with ever-escalating bandwidth and performance demands, are driving a rapid, industry-wide shift to all packet-based network services. To further complicate the picture, the convergence of wire line, wireless, mobile and Internet network services requires that heterogeneous elements work together. The new network services converge fixed or landline and mobile voice, video and data delivery using a packet-based transport that enables flexible, scalable and cost-efficient services.

This next phase of network evolution needs to happen quickly; service providers are already offering new services that challenge current network service capabilities. The explosion of Web-based applications, the use of Service Oriented Architectures (SOAs) and concepts such as software as a service (SaaS) raise the mission critical nature of applications to the point where reliability and availability are important factors for business success. This contrasts with the dependability of today's legacy Public Switched Telephone Network (PSTN) which has been realized through many years of rigorous engineering discipline, costly development efforts and controlled field maturity.

As communications equipment is deployed into packet-based multiservice networks, the dependability and availability of services must not be compromised. Users will expect new, innovative distributed services to be delivered on demand and without interruption. The only way all these demands of the emerging communications environment can be rapidly met is through the widespread adoption of industry standards. An important and necessary part of such standards is an open, standardized interface to an availability management framework.

A Paradigm Shift

The transformation of networks – the migration from discrete-service based architectures to converged networks that are IP-based, service oriented and transport agnostic – is not lost on those who supply telecom equipment to the network operators. In the last decade there has been a continued paradigm shift from a vertical to a horizontal industry model as equipment manufacturers must build communications equipment, and enterprises must develop applications, that achieve the highest possible levels of availability and dependability while contending with ever-shorter

development cycles and constant pressure to reduce development costs. Service providers, in turn, rapidly deploy new services and vouch for their availability and integrity in order to successfully compete for users and strive to meet customer service level agreements.

And this shift is not unique to the telecommunications market, as it is evident in a variety of different markets such as aerospace & defense (A&D), manufacturing, etc., where the providers of highly available systems and services are increasingly looking to the industry to supply building blocks they traditionally have designed and developed internally. The aerospace and defense industry, particularly in the U.S., is providing a strong incentive to their primes – their suppliers and system integrators – to include as much of the commercial off-the-shelf (COTS) building blocks as possible. Arguably, the A&D industry invented the term *COTS*. Similar trends are also merging in manufacturing and financial markets.

What is catalyzing this shift is the emergence of key standards – published by a set of special interest groups (SIGs) – that are creating clear delineation between various functional layers of a highly available system. This standardization of functional layers – hardware, operating system, middleware and application services – is greatly facilitating the ability for systems developers to put together highly available deployment ready systems using COTS building blocks [Figure 1]. The emergence of multiple COTS suppliers for each of the building blocks is helping create a viable and vibrant ecosystem that provides compelling alternatives for the systems developers to build systems leveraging a strong ecosystem of COTS component suppliers. As a result, system developers can focus their precious, often shrinking, resources on things that differentiate them from the competition – applications and services.

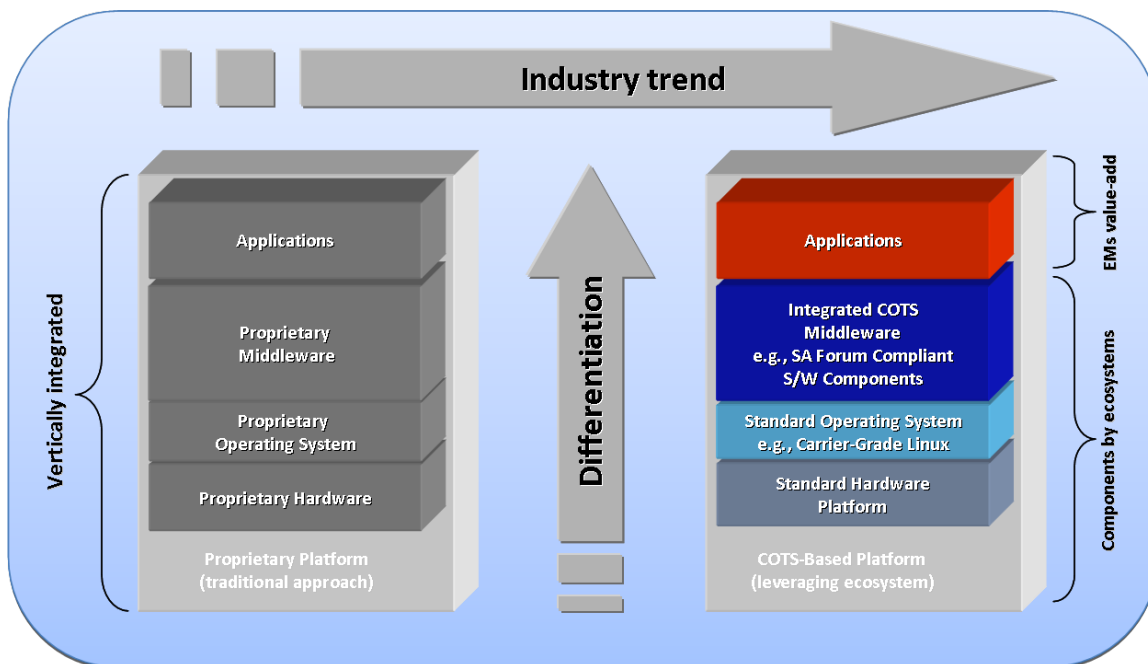


Figure 1. Key standards are catalyzing a move from proprietary to standards-based systems

One of the key requirements of a system worthy of deployment is that it has to be *highly available* – i.e., it must provide uninterrupted service even in the event of hardware or software failure. Historically NEPs have designed and built such systems from the ground up using the specialized, in-house expertise that they have developed and nurtured over decades. Many systems providers have invested a significant amount of time and resources in developing software services, often referred to as *high availability middleware*, essential to building platforms and systems that provide

service availability approaching five-nines or better (see Appendix A). Recently however, such middleware is becoming the focus of various Independent Software Vendors and open source projects that are providing service availability products that package a collection of key high availability and management services that can be acquired off-the-shelf and employed in a horizontal system design and implementation.

This proliferation has been made possible by the publication and adoption of application programming interface specifications (API) published by the SA Forum (www.saforum.org). The SA Forum provides a rich set of API specifications that address several areas: a hardware abstraction layer termed *Hardware Platform Interface* (HPI) and an application abstraction layer called the *Application Interface Specification* (AIS). In concert these specifications, when implemented, allow for portability and management of service availability middleware as well as applications that comply with them.

The Service Availability Forum™

The Service Availability Forum™ (SA Forum) is a consortium of industry-leading communications and computing companies working together to develop and publish high availability and related management software interface specifications. Since its inception in 2001, the SA Forum has focused its efforts on producing key specifications to address the requirements of availability, reliability and dependability for a broad range of applications. To date, the SA Forum has released the Hardware Platform Interface (HPI) and the Application Interface Specification (AIS). As implementations of these specifications are increasingly accepted in the marketplace, the SA Forum is accelerating its effort to educate the communications and computing industries in general on how to develop applications that can achieve high service availability.

Individuals from member companies enjoy exposure to peers, leaders, and best-in-class technology. Joining and engaging in the SA Forum has significant benefits that can enhance a company's stature and competitiveness in the industry. (See Appendix B for Membership and Benefits)

This paper explains the context and need for widespread adoption of a standard Service Availability interface, describes the Service Availability Forum solution, and identifies the benefits that an open, standard Service Availability interface will provide.

The Service Availability Forum Specifications

When implemented, the SA Forum specifications provide the following key benefits.

- The ability to discover, monitor and manage the hardware resources in the underlying platform in a manner that is hardware independent. This is accomplished through the implementation of services specified by the HPI. This results in the portability of the HPI-compliant service availability middleware across different hardware platforms that provide HPI services.
- Portability of AIS-compliant applications across multiple middleware implementations that provide services accessible via the interfaces specified by AIS
- Standardized access to management capabilities of HPI and AIS services
- The ability to perform in-service upgrade of various software components running on the system

An abstract view of a highly available system that utilizes SA Forum specifications is depicted in Figure 2.

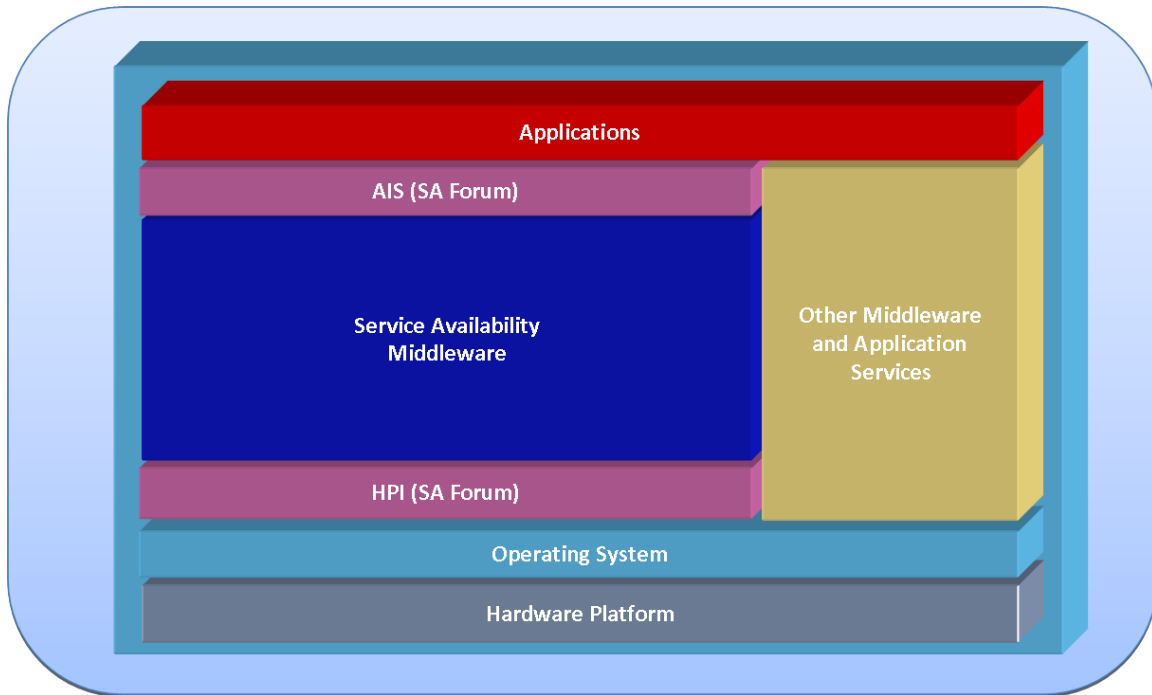


Figure 2. A high level view of a highly available system that utilizes SA Forum interfaces and COTS components

The SA Forum continues to enhance these specifications and periodically releases updated versions. The latest release includes interface specifications for a comprehensive set of services defined by HPI and AIS. A brief overview of these specifications is included here.

The Hardware Platform Interface Specification (HPI)

The HPI capabilities can be categorized in three broad areas:

- **Discovery** allows the capability to automatically discover and enumerate various hardware resources present within a hardware platform along with their management capabilities.
- **Monitoring** allows the capability to query various hardware resources to obtain their current health status, performance, etc.
- **Management** capabilities allow for appropriate manipulation of the state of a variety of management instruments such as sensors, controls, displays, annunciators, etc.

Typically, services specified by the HPI are implemented as libraries in the hardware platform. These services serve up the hardware information in a standard way so that a user application does not have to be aware of the specifics of the underlying hardware.

The Application Interface Specification (AIS)

AIS specifies a comprehensive set of functionality which, when implemented in the form of service availability middleware, provides a sophisticated high availability framework and associated services that can be used to develop highly available applications. Broadly speaking the AIS addresses a number of key dimensions of service availability middleware:

- An Availability Management Framework (AMF) defines the essential capability that coordinates redundant resources within a cluster to eliminate potential single points of failure. It provides a consistent view of one logical system that comprises a number of nodes, each of which hosts various resources in a distributed computing environment.

These resources are treated as managed objects or components and AMF manages their workload assignments dynamically as conditions may change in the system. The AMF assigns workloads/services to the redundant components to provide high availability.

- The managed objects for the whole system are represented in an information system model kept within the system. The Information Model Management (IMM) defines such a consistent information system model, and provides a cluster-wide mechanism to access, manage and control these managed objects.
- A broad set of APIs are specified that define core services required to implement a highly available system that ensures service availability even in the presence of hardware or software failures.
- The Software Management Framework (SMF) enables system developers to create capabilities to manage various software components within the system. Used in conjunction with AMF, SMF provides the ability to perform in-service software upgrade without impacting service availability.

A Service Availability Platform

Consider the anatomy of a highly available and manageable application-ready platform [Figure. 3]. The diagram reveals the key categories of services needed to build such a system.

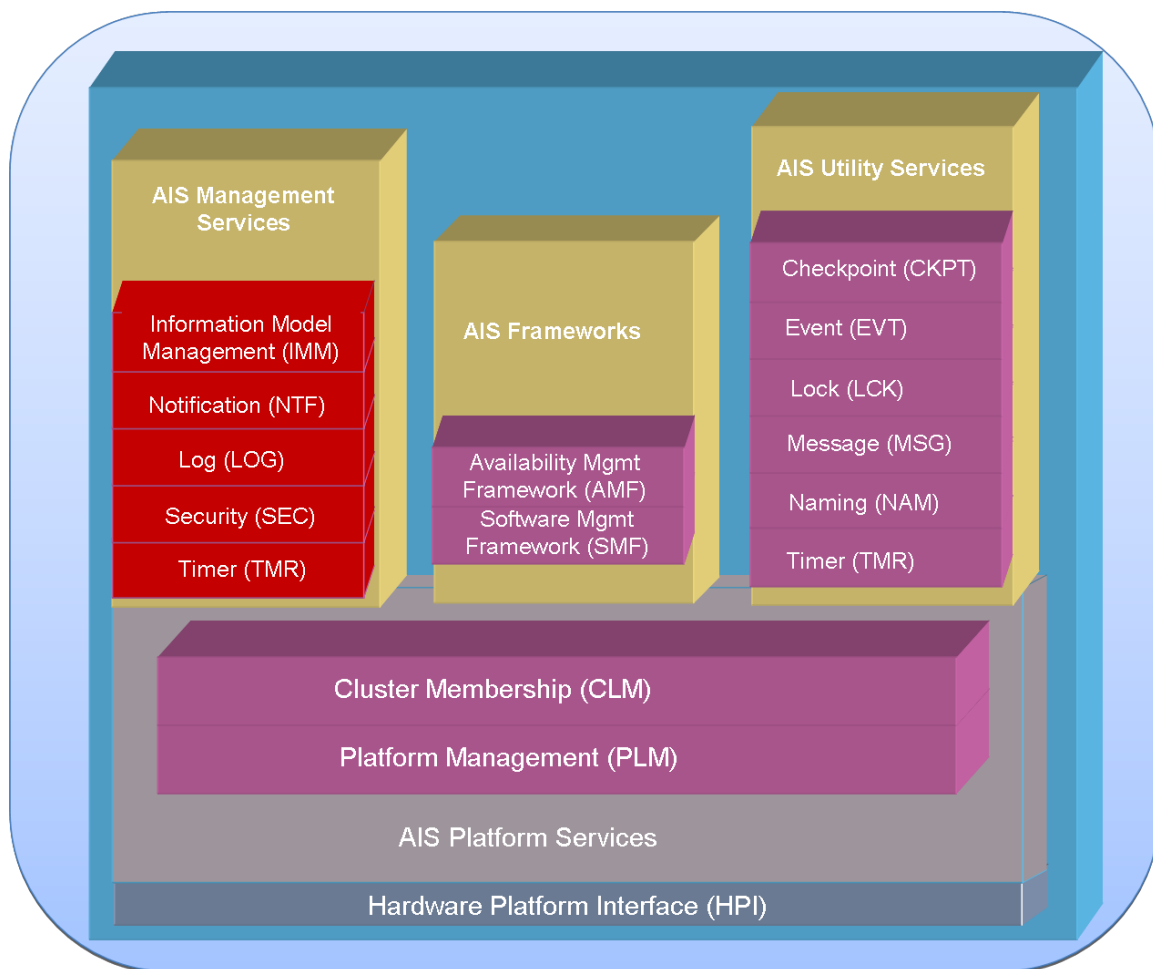


Figure 3. Essential services that are required to put together an application ready platform

AIS platform services

These services provide a higher-level abstraction of the hardware platform and operating systems to the other AIS services and applications:

- Platform Management Service: It provides a convenient abstraction for controlling configured hardware and low level software (HW/OS) resources. It defines APIs for monitoring state changes in application-defined groups of configured resources
- Cluster Membership Service: It provides a consistent view of the healthy nodes in a cluster.

AIS management services

These services provide the basic standard management interfaces that should be used by the implementation of all AIS services and applications.

- Information Model Management Service: It provides APIs for defining, obtaining, manipulating and exposing configuration and runtime management information as well as for invoking administrative commands on managed objects.
- Notification Service: It provides data structures and APIs for notifying alarms, state changes, object life cycle changes, attribute value changes, security alarms and other particular events (notifications of miscellaneous type).
- Log Service: It allows for cluster-wide logging of alarms, notifications, system messages and application-defined log streams.
- Security Service: It provides access security functions for the AIS services and HPI.

AIS frameworks

- Availability Management Framework: It provides functions for availability management of applications and middleware.
- Software Management Framework: It is used for managing middleware and application software during hardware, operating system, middleware and application upgrades while taking accountability into account.

Management interfaces

Management interfaces are used to provision, monitor, manage, troubleshoot and upgrade the entire system. Ecosystem choices vary, ranging from loose integration, where one interface is used for each layer of the system stack, to tight integration, where a single, unified management interface (UMI) is used to manage all stack layers from one source. From an operator standpoint, the UMI approach is preferable in that it allows one common portal through which to access all of the elements of the system while eliminating the need for multiple portals that must be compared and synchronized constantly.

AIS utility services

These services provide some of the common interfaces required in highly available distributed systems, including:

- Checkpointing
- Event distribution
- Distributed locks
- Message passing
- Name lookup
- Local timer service

Benefits

The SA Forum interface specifications facilitate standardized interfaces that provide the following benefits:

Shortened development cycles

- The use of COTS components and adoption of a standard interface in a building-block approach simplifies design and makes the system easier to deploy and maintain
- A common interface fosters an efficient and simultaneous approach that enables success
- A standard interface enables smarter utilization of resources (i.e. ability to outsource mission-critical development work that is not a core competency, focus on innovation, etc.)

Reduced costs

Standards-based design reduces the costs of developing, provisioning and maintaining infrastructure equipment. In particular, it addresses difficulty and/or expense associated with the following:

- Developing and testing on vertically integrated, proprietary equipment
- Porting of existing software and applications
- Deployment of proprietary equipment and systems
- Training and education of both technical and operational people
- Maintaining proprietary, vertically integrated systems

Increased design flexibility and risk reduction

The standards-based approach and an open, building-block system architecture enables equipment manufacturers to mix and match best-in-class or most-suitable components from a variety of competing vendors. These off-the-shelf products include hardware platforms, service availability middleware, operating systems and applications. Using commercial off-the-shelf components also allows design flexibility and frees up equipment manufacturers and application developers to focus on core competencies. In addition, a standard interface and open architecture enable manufacturers and system integrators architectural flexibility to more easily adapt to changing technologies over time.

Technological innovation

Widespread use of a multi-vendor, standards-based approach to system design helps promote competition and stimulate technical innovation. An open, standard service availability interface will spur innovation in three ways: 1) providing design flexibility, 2) simplifying the work of application developers and 3) allowing developers to focus on core competencies and to create more and better, revenue-generating applications. End users benefit from continuing improvement of the services delivered.

Increased benefits to the value chain

Each element of the communication infrastructure ecosystem will benefit from the widespread adoption of Service Availability standards.

- **Component and Board Manufacturers** will increase revenue and profits by providing the SA Forum-compliant platform solutions that offer quick and seamless platform integration into multiservice networks.
- **Systems and Applications Providers** (communications equipment manufacturers and software vendors) will decrease costs and improve time-to-market by moving from proprietary high availability solutions to open, industry-standard interfaces and COTS implementations.
- **Service Providers and Operators** can win and keep customers by ensuring that their revenue-generating applications and services are continuously available. Standards improve competitiveness by enabling fast integration and deployment of new revenue generating services. Furthermore, the standards will facilitate their ability to meet or exceed service level agreements with their customers, and to turn service availability benefits into a key differentiator in the marketplace. Standards and the open competition of the ecosystem model will also accelerate innovation of applications, enabling service providers to offer new, innovative services more rapidly.
- **End-users and IT Customers** will have more choices of new, dependable services. SA Forum-compliant platforms and applications will ensure that their mission critical, revenue-generating applications and services are continuously available and dependable.

Adoption

HPI was the first application programming interface specification released by the SA Forum. Through multiple revisions and significant industry feedback, it has evolved significantly and is now a significant market success. With numerous hardware suppliers providing HPI services as an integral part of their platforms, this specification has quickly become the standard of choice by the industry for providing platform management capabilities. The recent introduction of PLM has significantly enhanced the usefulness of this interface by bridging the gap between the AIS and HPI system views. AIS has also evolved, and commercial implementations of the AIS-compliant middleware services and applications that are built on such middleware are beginning to appear in the market. Several major systems and application providers are getting ready to commercially deploy their systems that employ AIS-compliant middleware services and applications that utilize such services. It is expected that AIS will see commercial success similar to HPI in the foreseeable future. Adoption of SA Forum specifications is emerging in other markets such as aerospace and defense, manufacturing, and others, where HPI and AIS are being used to develop and deploy systems for mission-critical applications.

A Last Word

The SA Forum work is transforming the way providers of highly available systems develop and commercialize platforms, applications and services. HPI enables not only sophisticated standards-based platform management capabilities, but also significantly reduces the R&D cost involved in developing and integrating such functionality. By providing a rich abstraction layer to the application and services developers, AIS shields the designers from the details of underlying middleware enabling them to develop portable applications. The net result is reduction in R&D and maintenance costs, more predictable schedules, and an increased velocity to revenue. The fact that these benefits are real and significant is evident in the increasing adoption of the SA Forum specifications by a variety of major industry players from different market segments.

The Service Availability Forum is committed to working with the industry to continue to assist in accelerating the adoption of the COTS business model by providing interface specifications that help create standard software building blocks for un-interrupted service availability. The SA Forum

is going one step further in ensuring that there are ways to demonstrate that applications and services authored using its specification are portable, if not transportable, across multiple implementations of underlying service availability middleware.

Resources

The Service Availability Forum also provides education material that delivers Web-based tutorials to those considering implementations based on its specifications. Please visit www.saforum.org for more information.

Appendix A

What is Service Availability?

While high availability is a measurement of system uptime of five-nines (99.999%) or better, it does not inherently account for the end-user's experience of continuous service delivery. A highly available system is not necessarily highly reliable. The five-nines metric could be met, for example, despite a hardware fault that is detected and repaired by the service availability middleware before the user was affected. In this hypothetical example, the system could be described as highly available even though it has failures visible to the service consumer because of loss of state.

Below are some important terms and definitions to clarify the difference between high availability and service availability.

High Availability

High Availability (HA) is the measure of the probability that a service is available for use at any given instant. Current infrastructure systems are commonly described as highly available if they have an uptime of five-nines (99.999%) or better. This translates to less than five minutes of cumulative downtime per year. Availability is a function of system reliability and reparability, augmented and supported by redundancy.

Reliability

Reliability is a measure of continuous system uptime in the absence of failure. Highly reliable systems are those that have a high mean time between failures (MTBF). When they do fail, however, such systems can experience lengthy outages that compromise system availability metrics.

Reparability

Achieving high availability is also a factor of how quickly a system can be repaired when failures do occur. Reparability is the measure of how quickly a failed device or system can be restored to service, calculated as mean time to repair (MTTR). The less reliable a system component is, the greater its need to have low MTTR to be able to support overall system availability.

Redundancy

Redundancy augments the reparability of individual components by establishing a backup or standby. When there are multiple resources providing the same functionality, they can take over for one another in case of failure. The effectiveness of redundancy is a function of how quickly a backup component can be brought into service. There are various redundancy configurations (2N, N+M, N-way, etc.) and readiness states (cold, warm and hot switchover) to appropriately support uptime requirements.

Availability is measured using the following formula:

$$\text{Availability} = \text{MTBF} / (\text{MTBF} + \text{MTTR})$$

The key to ensuring high availability is either to provide very reliable elements (very high MTBF), to provide elements that can recover from failure very rapidly (very low MTTR), or both. The higher the availability supplied by system reliability and reparability, the less redundancy support is required to maintain availability.

Service Continuity

End-users and service providers require a higher standard than high availability. They do not care about the state of the hardware; what they want is continuously available services delivered without disruption. Service continuity is the uninterrupted delivery of services from the end-user's

perspective. The service availability not only requires systems that are highly available, it carries the caveat that, in order to provide continuity of service to the end-user, the customer data and session state must be preserved across fault scenarios. While high availability refers to the state of a system, service continuity refers to specific instances of service on a system. Preservation of data and state across failure/switchover scenarios is a challenging and complex requirement, but it is key to providing expected levels of service continuity to the end-user 99.999% of the time. This requires, among other things, checkpointing of application and other software state to the standby component, and availability management that can provide rapid, stateful software failover as required. The service availability solution is measured by the availability and dependability of the services provided by the application(s) running on a system—in other words, the MTBF *of the service*. Continuous service uptime is dependent first upon reliable, repairable, redundant components and availability management that can provide high system uptime. Secondly, it is dependent upon availability management that also can orchestrate continuity of service, including preservation of application and other software state.

Appendix B

SA Forum Benefits of Membership

Benefits of Membership

Significant numbers of companies will use or implement products based on the SA Forum's specifications, but many institutions and companies also choose to take an active role in definition, development and marketing these standards. Member companies include hardware, firmware, operating system, middleware and application providers, as well as all tiers of network equipment providers, integrators, and academic institutions. The rewards and benefits of joining the SA Forum and actively participating are significant.

Leadership – The SA Forum is recognized as successful and influential in the communications industry. The SA Forum is also seeing traction in the enterprise computing and aerospace/defense industries. Institutions and companies that are listed as contributing members are viewed as industry leaders. In the tradition of the SA Forum, credit for specifications and the work of the forum is equally shared by all member companies. Therefore, as member companies and their products are evaluated, being a participant and not just an implementer demonstrates a higher level of commitment, knowledge, influence and credibility.

Influence – The output of the SA Forum can dramatically affect a product's architecture and implementation. Every contributing member has an equal opportunity to influence the requirements, design, implementation and evolution of each specification. Member companies can also propose and champion new specifications and supporting documentation. Influence within the marketing activities can also affect how member companies go to market and promote their specific SA Forum-based products.

Knowledge – The quality and success of the SA Forum can be attributed directly to the high level of expertise of the individuals and their ability to work as a diverse team. Due to the significant role the SA Forum plays in the industry, member companies engage with their best and brightest. Participants ranging from recent graduates to seasoned experts function in a collaborative environment where all ideas are explored. This results in exposure to leading edge innovation as well as proven and well established methodologies.

Insight – Demonstrated by the list of member companies, the SA Forum is at the center of activity in the industry. Members are in regular discussion with the other leading companies and institutions about industry trends and technology advancements. This insight can help forge winning market and product strategies.

Competitiveness – SA Forum specifications are publically available only after they are complete. However, every contributing company has full access to all specifications and work products of the SA Forum as they are being developed. This affords both users and implementers of SA Forum specifications a significant time advantage in developing product architectures, designs and implementations. Awareness of issue resolution and specification enhancements can also create an advantage in product support.

Membership

The SA Forum structure is simple and relatively flat with a board of directors overseeing two major work groups. The Technical Work Group is where specifications and other technical output is generated. The Technical Work Group is further broken down into work groups that focus on specific work products. The Marketing Work Group is responsible for awareness and adoption of the specifications. The work groups meet weekly on conference calls to guide ongoing specification and marketing work. Each quarter a face-to-face general meeting is held at a member's facility.

General meeting locations are rotated between Asia, North America and Europe. The general meetings typically include three days of intense general and work group meetings with an evening social event.

The SA Forum provides four levels of membership; Promoter, Contributor, Associate and Adopter. Promoters make up the board of directors and are selected according to industry position and demonstrated contributions over an extended period of time. The majority of members participate at the Contributor level. The Contributor level is where the work gets done. Contributors have full access to all work in progress and can choose to contribute to any work group. Member companies can have as many individuals as desired participate. However there is a one company, one vote rule. The Associate level is specifically for academic institutions and has the same privileges as Contributor members. The Adopter level simply allows a company to use the SA Forum Logo to acknowledge the use of the organization's specifications.

Joining and engaging in the SA Forum has significant benefits that can enhance a company's stature and competitiveness in the industry. Individuals from member companies enjoy exposure to peers, leaders and best-in-class technology.