

Service-Oriented Computing and Software Agents

Mike P. Papazoglou (mikep@uvt.nl)

*INFOLAB, Dept. of Information Systems and Management, Tilburg University
PO Box 90153, Tilburg 500 LE, The Netherlands*

Marco Aiello (aiellom@dit.unitn.it) and Paolo Giorgini
(pgiorgio@dit.unitn.it)

*Dep. of Information and Telecommunication Technologies, University of Trento
Via Sommarive 14, 38050 Trento, Italy*

Abstract. Service-Oriented Computing (SOC) is the computing paradigm that utilizes services as fundamental elements for developing applications/solutions. To build the service model, SOC relies on the Service Oriented Architecture (SOA), which is a way of reorganizing software applications and infrastructure into a set of interacting services. However, the basic SOA does not address overarching concerns such as management, service orchestration, service transaction management and coordination, security, and other concerns that apply to all components in a services architecture.

In this paper we introduce an Extended Service Oriented Architecture that provides separate tiers for composing and coordinating services and for managing services in an open marketplace by employing grid services and discuss how agent technology can be used to support the functions of the Extended SOA.

Keywords: Service-Oriented Computing, Service-Oriented Architectures, agent-oriented programming, service management, service orchestration

1. Introduction

Service-Oriented Computing (SOC) is the computing paradigm that utilizes services as fundamental elements for developing applications-solutions. Services are self-describing, platform-agnostic computational elements that support rapid, low-cost composition of distributed applications. Services perform functions, which can be anything from simple requests to complicated business processes. Services allow organizations to expose their core competencies programmatically over the Internet (or intra-net) using standard (XML-based) languages and protocols, and be implemented via a self-describing interface based on open standards.

Because services provide a uniform and ubiquitous information distributor for a wide range of computing devices (such as handheld computers, PDAs, cellular telephones, or appliances) and software platforms (e.g., UNIX or Windows), they constitute the next major step in distributed computing.

© 2004 Kluwer Academic Publishers. Printed in the Netherlands.

Services are offered by service providers - organizations that procure the service implementations, supply their service descriptions, and provide related technical and business support. Since services may be offered by different enterprises and communicate over the Internet, they provide a distributed computing infrastructure for both intra- and cross-enterprise application integration and collaboration. Clients of services can be other solutions or applications within an enterprise or clients outside the enterprise, whether these are external applications, processes or customers/users. To satisfy these requirements services should be *technology neutral* in that the invocation mechanisms (protocols, descriptions and discovery mechanisms) should comply with widely accepted standards. Services should also be *loosely coupled* as they must not require knowledge or any internal structures or conventions (context) at the client or service side. Finally services should support *location transparency*. Services should have their definitions and location information stored in a repository such as UDDI and be accessible by a variety of clients that can locate and invoke the services irrespective of their location.

Services come in two flavors: *simple* (stateful services) and *composite services* (stateless services). The unit of reuse with services is functionality that is in place and readily available and deployable as services that are capable of being managed to achieve the required level of service quality. Composite services involve assembling existing services that access and combine information and functions from possibly multiple service providers. For example, consider a collection of simple services that accomplish a specific business task, such as order tracking, order billing, and customer relationships management. An enterprise may offer a composite web service that composes these services together to create a distributed e-business application that provides customized ordering, customer support, and billing for a specialized product line (e.g., telecommunication equipment, medical insurance, etc). Accordingly, services help integrate applications that were not written with the intent to be easily integrated with other distributed applications and define architectures and techniques to build new functionality while integrating existing application functionality.

Service-based applications are developed as independent sets of interacting services offering well-defined interfaces to their potential users. This is achieved without the necessity for tight coupling of distributed applications between transacting partners, nor does it require pre-determined agreements to be put into place before the use of an offered service is allowed.

While the services encapsulate the business functionality, some form of inter-service infrastructure is required to facilitate service interac-

tions and communication. Different forms of this infrastructure are possible because services may be implemented on a single machine, distributed across a set of computers on a local area network, or distributed more widely across several wide area networks. A particularly interesting case is when the services use the Internet as the communication medium and open Internet-based standards. A *web service* is a specific kind of service that is identified by a URI and exhibits the following characteristics. It exposes its features programmatically over the Internet using standard languages and protocols, and it can be implemented via a self-describing interface based on open standards (e.g., XML interfaces which are published in network-based repositories).

Interactions of web-services occur as SOAP calls carrying XML data content and the service descriptions of the web-services are expressed using WSDL (Web Service Definition Language,) as the common (XML-based) standard. WSDL is used to publish a web service in terms of its ports (addresses implementing this service), port types (the abstract definition of operations and exchanges of messages), and bindings (the concrete definition of which packaging and transportation protocols such as SOAP are used to inter-connect two conversing end points). The UDDI (UDDI Technical White paper, 2001) standard is a directory service that contains service publications and enables web-service clients to locate candidate services and discover their details.

Web services share the characteristics of more general services, but they require special consideration as a result of using a public, insecure, low-fidelity mechanism for inter-service interactions.

This paper is organized as follows. In Section 2 we introduce the concept of software as a service, while in Section 3 we describe the basic service oriented architecture. Section 4 describes an extended service architecture materialized by grid services and Section 5 introduces the concept of service bus for open service marketplaces. Section 6 is devoted to agent-oriented technology and how it supports the service-oriented computing vision. Finally, Section 7 presents our conclusions.

2. A view of software as a service

The concept of software-as-a-service espoused by SOC is revolutionary and appeared first with the ASP (Applications Service Provider) software model. An ASP is a third party entity that deploys, hosts and manages access to a packaged application and delivers software-based services and solutions to customers across a wide area network from a central data center. Applications are delivered over networks on a

subscription or rental basis. In essence, ASPs were a way for companies to outsource some or all aspects of their information technology needs.

By providing a centrally hosted Intent application, the ASP takes primary responsibility for managing the software application on its infrastructure, using the Internet as the conduit between each customer and the primary software application. What this means for an enterprise is that the ASP maintains the application, the associated infrastructure, and the customer's data and ensures that the systems and data are available whenever needed.

Although the ASP model introduced the concept of software-as-a-service first, it suffered from several inherent limitations such as the inability to develop highly interactive applications, inability to provide complete customizable applications (Goepfert and Whalen, 2002). This resulted in monolithic architectures, highly fragile, customer-specific, non-reusable integration of applications based on tight coupling principles. Today we are in the midst of another significant development in the evolution of software-as-a-service architected for loosely-coupled asynchronous interactions on the basis of XML-based standards with intention to make access to and communications of applications over the Internet easier.

The SOC paradigm allows the software-as-a-service concept to expand to include the delivery of complex business processes and transactions as a service, while permitting applications to be constructed on the fly and services to be reused everywhere and by anybody. Perceiving the relative benefits of (web) services technology many ASPs are modifying their technical infrastructures and business models to be more akin to those of web service providers.

3. The basic service oriented architecture

To build integration-ready applications the service model relies on the service-oriented architecture (SOA). SOA is a way of reorganizing a portfolio of previously siloed software applications and support infrastructure into an interconnected set of services, each accessible through standard interfaces and messaging protocols. Once all the elements of an enterprise architecture are in place, existing and future applications can access these services as necessary without the need of convoluted point-to-point solutions based on inscrutable proprietary protocols. This architectural approach is particularly applicable when multiple applications running on varied technologies and platforms need to communicate with each other. In this way, enterprises can mix and match services to perform business transactions with minimal effort.

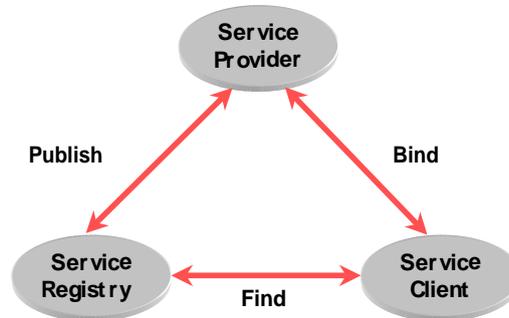


Figure 1. The basic Service Oriented Architecture.

SOA is a logical way of designing a software system to provide services to either end-user applications or other services distributed in a network through published and discoverable interfaces. The basic SOA defines an interaction between software agents as an exchange of messages between service requesters (clients) and service providers. Clients are software agents that request the execution of a service. Providers are software agents that provide the service. Agents can be simultaneously both service clients and providers. Providers are responsible for publishing a description of the service(s) they provide. Clients must be able to find the description(s) of the services they require and must be able to bind to them. We examine software agents in Section 6.

The basic SOA is not an architecture only about services, it is a relationship of three kinds of participants: the *service provider*, the *service discovery agency*, and the *service requestor (client)*. The interactions involve the *publish*, *find* and *bind* operations (Champion et al., 2002), see Figure 1. These roles and operations act upon the service artifacts: the service description and the service implementation. In a typical service-based scenario a service provider hosts a network accessible software module (an implementation of a given service). The service provider defines a service description of the service and publishes it to a client or service discovery agency through which a service description is published and made discoverable. The service requestor uses a find operation to retrieve the service description typically from a the discovery agency, i.e., a registry or repository like UDDI, and uses the service description to bind with the service provider and invoke the

service or interact with service implementation. Service provider and service requestor roles are logical constructs and a service may exhibit characteristics of both.

The fundamental logical view of a service in the basic SOA is that it is a service interface and implementation. A service is usually a business function implemented in software, wrapped with a formal documented interface that is well known and known where to be found not only by agents who designed the service but also by agents who do not know about how the service has been designed and yet want to access and use it. Services are intended to represent meaningful business functionality that can be assembled into larger and new configurations depending on the need of particular kinds of users.

The interface simply provides the mechanism by which services communicate with applications and other services. Technically, the service interface is the description of the signatures of a set of operations that are available to the service client for invocation. The service specification must explicitly describe all the interfaces that a client of this service expects as well as the service interfaces that must be provided by the environment into which the service is assembled/composed. As service interfaces of composed services are provided by other (possibly singular) services, the service specification serves as a means to define how a composite service interface can be related to the interfaces of the imported services and how it can be implemented out of imported service interfaces. This is shown in Figure 2. In this sense the service specification has a mission identical to a composition meta-model that provides a description of how the web-service interfaces interact with each other and how to define a new web-service interface ($\langle\text{PrortType}\rangle$) as a collection (assembly) of existing ones (imported $\langle\text{PrortType}\rangle$ s), see Figure 2. A service specification, thus, defines the encapsulation boundary of a service, and consequently determines the granularity of replaceability of web-service interface compositions. This is the only way to design services reliably using imported services without knowledge of their implementations. As service development requires that we deal with multiple imported service interfaces it is useful to introduce this stage the concept of a *service usage interface*. A service usage interface is simply the interface that the service exposes to its clients (Papazoglou and Yang, 2002). This means that the service usage interface is not different from the imported service interfaces in Figure 2, it is, however, the only interface viewed by a client application.

Figure 2 distinguishes between two broad aspects of services: *service deployment*, which we examined already, versus *service realization*. The service realization strategy involves choosing from an increasing diversity of different options for services, which may be mixed in various

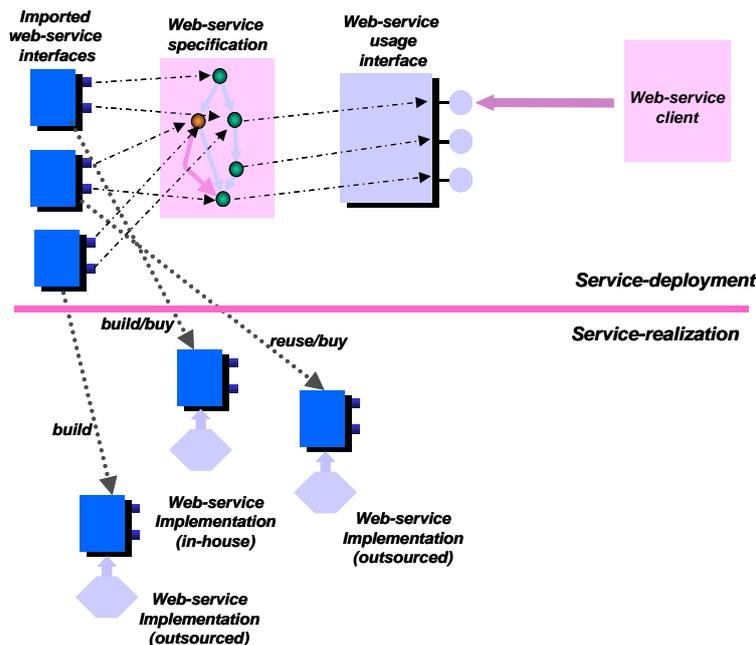


Figure 2. Service interfaces and implementation.

combinations including in house service design and implementation, purchasing/leasing/paying for services, outsourcing service design and implementation, and using wrappers and/or adapters to convert legacy system, COTS package, and ERP functionality into a service layer.

Service descriptions are used to advertise the service capabilities, interface, behavior, and quality. Publication of such information about available services (on a service registry) provides the necessary means for discovery, selection, binding, and composition of services. In particular, the service interface description publishes the service signature while the service capability description states the conceptual purpose and expected results of the service. The (expected) behavior of a service during its execution is described by its service behavior description (e.g., as a workflow process). Finally, the Quality of Service (QoS) description publishes important functional and non-functional service quality attributes, such as service metering and cost, performance metrics (response time, for instance), security attributes, (transactional) integrity, reliability, scalability, and availability.

Service implementation can also be very involved because in many occasions many organizations rely on single monolithic programs to represent the single service or service method implementation. But very often in order to fulfil the functions of a service multiple programs are involved, e.g., programs that belong to multiple applications of new

programs that belong to multiple applications. Application composition and integration very often is involved in fulfilling the service. At the logical level of the service we do not pay any attention to this. All we need to know is that there is a business function implemented in software somehow and this is the interface to it. At development time we care, however, how the service is implemented. More specifically, what are the methods and the internal construction of the implementation.

The service in the basic SOA is designed in such a way that it can be invoked by various service clients and is logically decoupled from any service caller (*loose coupling*). Services can be reused and one does not have to look inside the service to understand what it does. There are no assumptions of any kind in the service as to what kind of service consumer is using and for what purpose and in what context. In SOA the service is not coupled with its callers, in fact it knows nothing about them. However, the service callers are very much coupled with the service as they know what the services are what they call and what they can accomplish.

4. Grid services and the extended SOA

The basic SOA does not address overarching concerns such as management, service orchestration, service transaction management and coordination, security, and other concerns that apply to all components in a services architecture. Such concerns are addressed by the extended SOA (ESOA) (Papazoglou and Georgakopoulos, 2003) that is depicted in Figure 3. This layered architecture utilizes the basic SOA constructs as its bottom layer.

The *service composition layer* in the ESOA encompasses necessary roles and functionality for the consolidation of multiple services into a single composite service. Resulting composite services may be used by *service aggregators* as components (i.e., basic services) in further service compositions or may be utilized as applications/solutions by service clients.

Service aggregators thus become service providers by publishing the service descriptions of the composite service they create. A service aggregator is a service provider that consolidates services that are provided by other service providers into a distinct value added service. Service aggregators develop specifications and/or code that permit the composite service to perform functions that include the following:

- *Coordination*: controls the execution of the component services, and manages dataflow among them and to the output of the com-

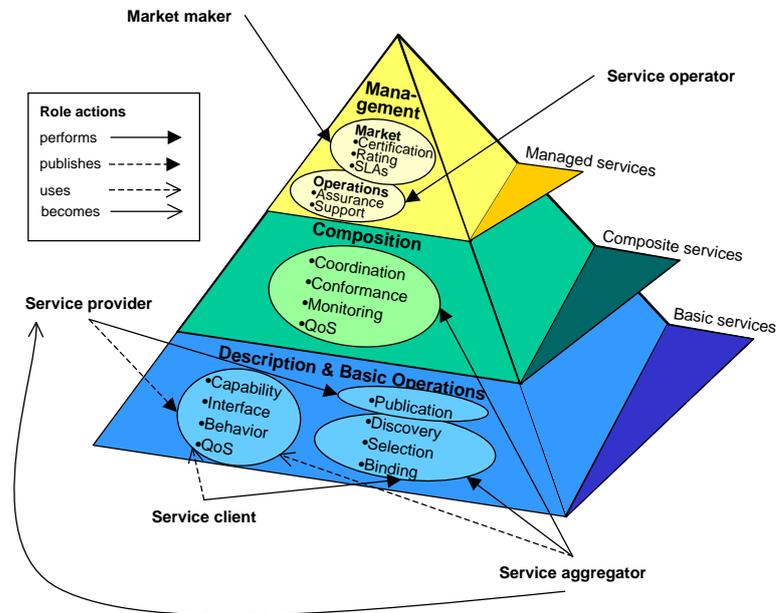


Figure 3. The Extended Service Oriented Architecture.

ponent service (e.g., by specifying workflow processes and using a workflow engine for run-time control of service execution).

- *Monitoring*: allows subscribing to events or information produced by the component services, and publish higher-level composite events (e.g., by filtering, summarizing, and correlating events).
- *Conformance*: ensures the integrity of the composite service by matching its parameter types with those of its components, imposes constraints on the component services (e.g., to ensure enforcement of business rules), and performs data fusion activities.
- *QoS composition*: leverages, aggregates, and bundles the component’s QoS to derive the composite QoS, including the composite service’s overall cost, performance, security, authentication, privacy, (transactional) integrity, reliability, scalability, and availability.

- *Policy enforcement*: web-service capabilities and requirements can be expressed in terms of policies. For example, knowing that a service supports a web-services security standard such as WS-Security is not enough information to enable successful composition. The client needs to know if the service actually requires WS-Security, what kind of security tokens it is capable of processing, and which one it prefers. The client must also determine if the service requires signed messages. And if so, it must determine what token type must be used for the digital signatures. And finally, the client must determine when to encrypt the messages, which algorithm to use, and how to exchange a shared key with the service. Trying to orchestrate with a service without understanding these details will lead to erroneous results.

The recently proposed standard Business Process Execution Language for web services (BPEL) (Curbera et al., 2002) is an XML-based effort to address the definition of a new web service in terms of compositions of existing services. A BPEL process is defined “in the abstract” by referencing and inter-linking portTypes specified in the WSDL definitions of the web services involved in a process.

Managing critical web-service based applications requires in-depth administration capabilities and integration across a diverse, distributed environment. For instance, any downtime of key e-business systems has a negative impact on businesses and cannot be tolerated. To counter such a situation, enterprises need to constantly monitor the health of their applications. The performance should be in tune, at all times and under all load conditions. Web-service based application management is an indispensable element of the ESOA that includes performance management and business/application specific management. This requires that a critical characteristic be realized: that services can be managed. Service management includes many interrelated functions. The most typical functions include:

1. *Deployment*: The web-services support environment should allow the service to be redeployed (moved) around the network for performance, redundancy for availability, or other reasons.
2. *Metrics*: The web-services support environment should expose key operational metrics of a web-service, at the operation level, including such metrics as response time and throughput. In addition it should allow web-services to be audited.
3. *Dynamic rerouting*: The web-services support environment should support dynamic rerouting for fail over or load balancing.

4. *Life cycle/State management*: The web-services support environment should expose the current state of a service and permit life-cycle management including the ability to start and stop a service.
5. *Configuration*: The web-services support environment should support the ability to make specific configuration changes to a deployed web-service.
6. *Change management and notification*: The web-services support environment should support the description of versions of web-services and notification of a change or impending change to the service interface or implementation.
7. *Extensibility*: The web-services support environment should be extensible and must permit discovery of supported management functionality in a given instantiation.
8. *Maintenance*: The web-services support environment should allow for the management and correlation of new versions of the service.

Web services manageability could be defined as the functionality required for discovering the existence, availability, performance, health, patterns of usage, extensibility, as well as the control and configuration, life-cycle support and maintenance of a web-service or business process within the context of the extended services architecture. This definition implies that web services can be managed using web-services technologies. In particular, it suggests a manageability model that applies to both web-services and business processes in terms of manageability topics, (identification, configuration, state, metrics, and relationships) and the aspects (properties, operations and events) used to define them. In fact, these abstract concepts apply to understanding and describing the manageability information and behaviour of any resource, including business processes and web-services.

To manage critical applications/solutions and specific markets, ESOA provides managed services in the service management layer depicted at the top of the ESOA pyramid. The ESOA managed services are divided in two complementary categories:

- *Service operations management* that can be used to manage the service platform, the deployment of services and the applications and, in particular, monitor the correctness and overall functionality of aggregated/orchestrated services.
- *Open service marketplace management* that supports typical supply chain functions and by providing a comprehensive range of

services supporting industry-trade, including services that provide business transaction negotiation and facilitation, financial settlement, service certification and quality assurance, rating services, service metrics, and so on.

The ESOA's service operations management functionality is aimed at supporting critical applications that require enterprises to manage the service platform, the deployment of services and the applications. ESOA's service operations management typically gathers information about the managed service platform, web-services and business processes and managed resource status and performance, and supporting specific management tasks (e.g., root cause failure analysis, SLA monitoring and reporting, service deployment, and life cycle management and capacity planning). Operations management functionality may provide detailed application performance statistics that support assessment of the application effectiveness, permit complete visibility into individual business processes and transactions, guarantee consistency of service compositions, and deliver application status notifications when a particular activity is completed or when a decision condition is reached. We refer to the organization responsible for performing such operation management functions as the service operator. Depending on the application requirements a service operator may be a service client or aggregator.

In the context of service operations management it is increasingly important for management to define and support active capabilities versus traditional passive capabilities. For example, rather than merely raising an alert when a given web-service is unable to meet the performance requirements of a given service-level agreement, the management framework should be able to take corrective action. This action could take the form of rerouting requests to a backup service that is less heavily loaded, or provisioning a new application server with an instance of the software providing the service if no backup is currently running and available.

Service operations management should also provide global visibility of running processes, comparable to that provided by Business Process Management (BPM). BPM promises the ability to monitor both the state of any single process instance and all instances in the aggregate, using present real-time metrics that translate actual process activity into key performance indicators (KPIs). Management visibility is expressed in the form of real-time and historical reports, and in triggered actions. For example, deviations from KPI target values, such as the percent of requests fulfilled within the limits specified by a service level agreement, might trigger an alert and an escalation procedure.

Considerations need also be made for modelling the scope in which a given service is being leveraged (individual, composite, part of a long-running business process, and so on). Thus, in addition to the above concerns, which relate to individual business processes or services, in order to successfully compose web-services processes), one must fully understand the service's WSDL contract along with any additional requirements, capabilities, and preferences (also referred to as policies). For example, knowing that a service supports a web-services security standard such as WS-Security is not enough information to enable successful composition. The client needs to know if the service actually requires WS-Security, what kind of security tokens it is capable of processing, and which one it prefers. The client must also determine if the service requires signed messages. And if so, it must determine what token type must be used for the digital signatures. And finally, the client must determine when to encrypt the messages, which algorithm to use, and how to exchange a shared key with the service. Trying to orchestrate with a service without understanding these details will lead to erroneous results. Such concerns are addressed by the service operations management. Service operations management is a critical function that can be used to monitor the correctness and overall functionality of aggregated/orchestrated services thus avoiding a severe risk of service errors. In this way one can avoid typical errors that may occur when individual service-level agreements (SLAs) are not properly matched. Proper management and monitoring provides a strong mitigation of risks due to mismatches of aggregated SLAs, since the operations management level allows to check the correctness, consistency and adequacy of the mappings between the input and output service operations and aggregate services in a service composition.

Another aim of ESOA's service management layer is to provide support for open service marketplaces. Currently, there exist several vertical industry marketplaces, such as those for the semiconductor, automotive, travel, and financial services industries. Open service marketplaces operate much in the same way like vertical marketplaces, however, they are open. Their purpose is to create opportunities for buyers and sellers to meet and conduct business electronically, or aggregate service supply/demand by offering added value services and grouping buying power (just like a co-op). The scope of such a service marketplace would be limited only by the ability of enterprises to make their offerings visible to other enterprises and establish industry specific protocols by which to conduct business. Open service marketplaces typically support supply chain management by providing to their members a unified view of products and services, standard business terminology, and detailed business process descriptions. In addition, service

marketplaces must offer a comprehensive range of services supporting industry-trade, including services that provide business transaction negotiation and facilitation, financial settlement, service certification and quality assurance, rating services, service metrics such as number of current service requesters, average turn around time, and manage the negotiation and enforcement of SLAs. ESOA's service management layer includes market management functionality (as illustrated in Figure 3) that is aimed to support these marketplace functions. The marketplace is created and maintained by a *market maker* (a consortium of organizations) that brings the suppliers and vendors together. The market maker assumes the responsibility of marketplace administration and performs maintenance tasks to ensure the administration is open for business and, in general, provides facilities for the design and delivery of an integrated service that meets specific business needs and conforms to industry standards.

The ESOA service management functions can benefit from grid computing as it targets manageability. Service grids constitute a key component of the distributed services management as the scope of services expands beyond the boundaries of a single enterprise to encompass a broad range of business partners, as is the case in open marketplaces. For this purpose grid services can be used to provide the functionality of the ESOA's service management layer (Foster et al., 2002; Tuecke et al., 2002). Grid services used in the ESOA's service management layer to provide an enabling infrastructure for systems and applications that require the integration and management of services with the context of dynamic virtual marketplaces. Grid services provide the possibility to achieve end-to-end qualities of service and address critical application and system management concerns.

5. The service grid bus

Grid services used in the ESOA's service management layer to provide an enabling infrastructure for systems and applications that require the integration and management of services with the context of dynamic virtual marketplaces. Grid services provide the possibility to achieve end-to-end qualities of service and address critical application and system management concerns. To this end grid services provide the grid infrastructure over which services interact, aggregate, and coordinate through a distinct architectural tier. This infrastructure is called the *service grid bus*. The service grid bus (SGB) provides a high-level abstraction architecture and management facilities to allow services (within an open service marketplace) to function as an integral unit and

collaborate with other services. The SGB architecture provides facilities for registration, discovery, selection/routing, business rules, filtering, routing, aggregation, fail-over, and topological mapping of service instances. The business rules govern the SGB's automatic processing for incoming service requests over aggregated service instances.

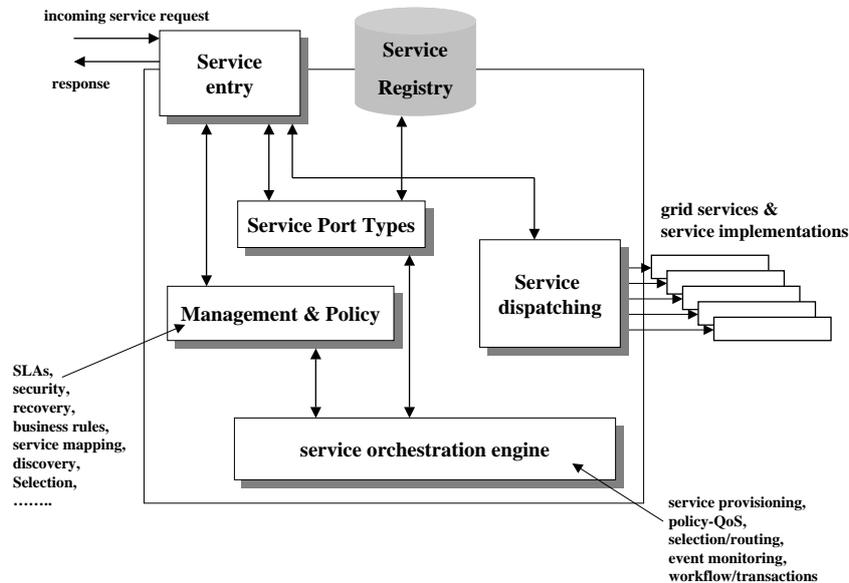


Figure 4. The service grid bus.

The service bus is a logical construct that cares very little where or on what platform a service provider runs. A user interface, designed to exactly match the business requirements may consume the services provided by the service bus, leaving the enterprise a free hand to choose the most efficient service provider.

An SGB is designed to provide a single service connectivity and a management tier that addresses the following concerns:

- Reach and robustness: The SGB allows to locate services anywhere in an open service marketplace and insulate them from connection failures, errors, and barriers such as firewalls, proxies, and caches. It guarantees that each service always sees an error-free connection to any other service. The application as a whole should be robust under transient or long-term failure of one or more service components.
- Policy and security management: Services need to describe their capabilities and requirements to their environment and potential

users. A collection of capabilities and requirements is referred to as a policy (Leymann, 2003). A policy may express such diverse characteristics as service transactional capabilities, security, response time, pricing, etc. For example, a policy of a service may specify that all interactions must be invoked under transaction protection, that incoming messages have to be encrypted, that outgoing messages will be signed, that responses may only be accepted within a specific time interval, etc. Finally, services must be restricted to authorized producers and consumers. The SGB enforces the security policy uniformly and universally across the entire marketplace.

- Development time and cost: The SGB provides the facilities that allow services to be easily aggregated into composite, higher-level services that match the factoring of an application. New services and clients will be added throughout the lifetime of the application. That process must be managed quickly, efficiently and cost effectively.
- Scalability and performance: The SGB must be able to perform well despite slow components and long, unpredictable network latencies.

The SGB lets service components interact over any network connection, handling all network errors, barriers, and transient or extended off-line conditions. It provides support for a business transaction model and support mechanisms for advanced transactional behavior of complex service-oriented business processes that span organizations (Papazoglou, 2003). The transaction model allows expressing unconventional atomicity criteria, e.g., payment atomicity, conversation atomicity, contract atomicity, and possesses the ability to express collaborative agreements and business conversation sequences that rely on transactional support. The model relies on a phased approach to business transactions so that all exchange of information between partners on the terms they could commit to, e.g., to fix price and quantity, are kept outside the “pure” transaction protocol. This results in enhancing flexibility and reducing latency and expensive transaction compensations and rollbacks in business interactions. The SGB’s asynchronous communications makes the application robust under transient service node failures, and allows transparent rerouting in case of prolonged failure.

The SGB provides standard, high-level services for security, management, service interaction, and aggregation, greatly accelerating application development and deployment. The SGB should also be consistent with emerging Web services workflow and transaction standards such as Web Services Transactions (F. Cabrera et al, 2002a), Web Services

Coordination (F. Cabrera et al, 2002b) and the Business Transaction Protocol (BTP) (OASIS Committee Specification, 2002). Finally, the SGB supports linear scalability and high performance by offering native support for asynchronous interaction, and allowing service endpoints to be managed for load balancing, workload distribution, and fail-over. The SGB itself must be scalable and capable of supporting peak loads.

Figure 4 describes a service sharing and aggregation SGB model for open service marketplaces. The SGB model allows services and the resources they use to be more easily shared by different constituencies within an open service marketplace. With its service grid foundation an open service marketplace provides the notion of a *business service grid* that automatically dispatches the best service available from a pool of dynamically assembled service providers in order to meet the user's need. Selection of a service is not just based on availability, but can also be based on QoS characteristics, as specified in SLAs and business arrangements. Selection of a service is performed automatically based on service policy, and the features provided by the SGB enable service providers to conceal the implementation complexity required to handle multiple client requests over heterogeneous environments.

The SGB invokes a service aggregation module to maintain its policy and states while serving external service requests. This module also performs selection and dispatching to find a service instance to execute a service request. The service aggregation module is also decomposed into modular functional units so that it is customizable to meet the special needs of diverse platforms, operations logic and so on.

One of the major benefits of introducing a distinct integration tier in the form of the SGB to implement the ESOA's service management layer is the ability to couple in value-added services that provide packaged solutions for common development needs. The SGB provides an asynchronous interaction model that lets users and applications initiate and monitor multiple tasks and data feeds simultaneously, and immediately alerts them when a transaction completes or a critical event occurs. In this regard, the SGB allows dynamic data feeds and transaction events to be delivered to the users or applications immediately when they occur. The SGB delivers to users applications that let them view multiple key performance indicators in real-time, collaborate with other users, and take action when critical events take place.

6. Software agents and the extended SOA

The shift towards the Service Oriented Computing paradigm not only involves a new way of conceptualizing complex distributed applications,

but also calls for more abstract software development and deployment technologies. Agent-oriented software engineering delivers such an abstraction (Giorgini et al., 2004).

To be qualified as an agent, a system is often required to have properties such as autonomy, social ability, reactivity, and proactivity. Other attributes which are sometimes requested (Wooldridge and Jennings, 1995) are mobility, veracity, rationality, and so on. The key feature which makes it possible to implement systems with the above properties is that programming is done at a very abstract level, more precisely, using a terminology introduced by Newell, at the *knowledge level* (Newell, 1982). Thus, we talk of mental states, of beliefs instead of machine states, of plans and actions instead of programs, of communication, negotiation and social ability instead of interaction and I/O functionalities, of goals, desires, and so on. Mental notions provide, at least in part, the software with the extra flexibility needed in order to deal with the intrinsic complexity of the applications mentioned in the first paragraph. The explicit representation and manipulation of goals and plans facilitates, for instance, a run-time “adjustment” of the system behavior needed to cope with unforeseen circumstances, or for more meaningful interactions with other human and software agents.

In the agent-oriented vision, software is built not by providing low level imperative lines of code to be followed sequentially, but rather by defining high-level goals to be achieved. In order for an agent to achieve its goals it must have a number of capabilities, most notably, an agent must have reasoning, communicating, and acting abilities.

We now focus on the three main agent capabilities in the context of web services (Section 3): *reasoning*, *communicating*, and *acting*.

Reasoning. Given a set of goals, knowledge about the world’s behavior, and information on the current status of the world, the agent must be able to reason to achieve the goals that have been delegated to him. The world is an electronic marketplace or a business grid, which is well structured, standardized and follows precise semantically defined business rules. Knowing a business process in the form of a BPEL description, a security or transactional requirement or a business rule in the form of a WS-Policy specification allows the agent to model precisely the world’s behavior and therefore to plan and act accordingly.

Communicating. Agents must be able to communicate one another and with services in order to cooperate, reach agreements and negotiate business interactions. This must happen via standardized languages such as those offered by XML-based web services

descriptions. For instance, agents may want to discover services offered in the world they populate by standard communication with UDDI registries, they may want to subscribe to events published via standard WS-Notification language (Graham et al., 2004).

Acting. To achieve the goals delegated to them agents must act (proactively and autonomously) in the environment they populate. In the web service scenario, acting implies invocation of web service functionalities. This is possible by invoking web service operations described in the standard WSDL language. In this manner, agents are able to accomplish the task assigned to them.

In general, the more the world in which the agent lives is explicitly semantically defined, the easier it is for the agent to reason, plan and, hence, achieve its goals. Web services give quite some structure to the world in which agents live as, for instance, they force a common language for interaction (e.g., WSDL to define basic web service operations) or even provide state-based descriptions of an application's behavior (e.g., a BPEL abstract process specification). However, this is not enough because we are only dealing with standardized syntactic descriptions with no semantic meaning attached to them. WSDL and BPEL specifications include no semantic attachments. This is potentially dangerous when specifying services in the context of processes. Consider, for instance, two semantically identical services providing, say, the same insurance functionality. These could end up with two completely different BPEL descriptions.

Two options are possible to overcome such a semantic gap between the web service world and the agent technology: add semantics to the web services or add knowledge and abilities to the agents. The semantic web service initiative (McIlraith et al., 2001) goes in the first direction, proposing to semantically tag all web service descriptions, while other proposals go in the second direction enabling agents to be designed for specific web service tasks (Li et al., 2003).

The latter option holds more promise to the world of service oriented computing as it is in line with the characteristics and requirements of agent technology that we mentioned in the outset of this section. These are also in line with previous research activities (Papazoglou, 2001) in the area of e-business where we identified various types of business agents based on business role they perform: application agents, personal agents, general business activity agents, information brokering agents, negotiation and contracting agents, system-level support agents, e.g., agents for interoperation, planning and scheduling agents, business transaction agents, and security agents.

In the context of service oriented computing we foresee that software agents could make an important contribution in connection with the roles undertaken in an ESOA. Agent technology could be used to provide two types of coarse agents: *agents as service providers* and *agents as service clients*. These two types of agents could be combined when a coarse agent behaves as a service aggregator (or market maker) in order to provide an added value services to other agents.

When considering agent technology in support of service aggregator agents we could expect that each coarse aggregator agent could comprise five distinct types of more fine-grained agents which provide the five paramount functions for service aggregators that we identified in Section 4. Specifically, we could identify the following types of agents:

- *Coordination agents*: these control the execution of single services and interact with them to achieve high-level complex business goals that have been delegated to them.
- *Monitoring agents*: these sense the world by subscribing to events of individual services and react to them by publishing higher-level events or communicating with other agents.
- *Conformance agents*: these achieve complex goals which may be conformance goals. A business rule may be formalized as a goal for the agent which will be enforced as the agent is composing a set of services.
- *QoS composition agents*: the goal of an agent may contain quality of service QoS specifications that the agent will achieve in composing the services. For instance an agent may choose only services that guarantee a specific response time in order to guarantee a specific total response time for the composed service.
- *Policy enforcement agents*: as for conformance and quality of service, policy constraints may be embedded in agent goals, therefore, an agent may enforce policies of aggregated services when composing them into an added value service.

In a similar manner to aggregator agents, market maker agents are also coarse agents comprising different types of more specialized fine-grained agents providing the service management functions required for managing service operations, open service marketplace and the business service grid functions (see Section 5) required for the SGB. Here we could have agents such as deployment agents, service selection agents (based QoS criteria such as price, performance, availability), rerouting

agents, life-cycle management agents, configuration/versioning agents, change management agents and so on.

7. Conclusions

In this paper we introduced the concepts behind Service Oriented Computing and explained how the basic Service Oriented Architecture helps deliver service-based applications. We argued that in order to provide the advanced functionality needed to deliver sophisticated e-business applications an Extended Service Oriented Architecture is necessary. This architecture includes a service composition tier to offer necessary roles and functionality for the consolidation of multiple services into a single composite service. It also provides a tier for service operations management that can be used to monitor the correctness and overall functionality of aggregated/orchestrated services and support for open service marketplaces. We explained how grid services can be used to implement the service management tier of the Extended Service Oriented Architecture by means of the service grid bus. Finally, we have shown agent-oriented technology to be an enabling technology for achieving the Service Oriented Computing paradigm.

References

- Champion, M., C. Ferris, E. Newcomer, and D. Orchard: 2002, 'Web Services Architecture'. *W3C Working Draft*. <http://www.w3.org/TR/ws-arch/>.
- Curbera, F., Y. Golland, J. Klein, F. Leyman, D. Roller, S. Thatte, and S. Weerawarana: 2002, 'Business Process Execution Language for Web Services 1.0, (BPEL4WS)'. <http://www.ibm.com/developerworks/library/ws-bpel/>.
- F. Cabrera et al: 2002a, 'Web Services Coordination (WS-Coordination)'. <http://www.ibm.com/developerworks/library/ws-coor/>.
- F. Cabrera et al: 2002b, 'Web Services Transaction (WS-Transaction)'. <http://www.ibm.com/developerworks/library/ws-transpec/>.
- Foster, I., C. Kesselman, J. M. Nick, and S. Tuecke: 2002, 'Grid Services for Distributed System Integration'. *IEEE Computer* **35**(6).
- Giorgini, P., P. Jörg, P. Müller, and J. Odell (eds.): 2004, *Agent-Oriented Software Engineering IV*, Vol. 2935 of *Lecture Notes in Computer Science (LNCS)*. Springer-Verlag.
- Goepfert, J. and M. Whalen: 2002, 'An Evolutionary View of Software as a Service'. DC White paper, <http://www.idc.com>.
- Graham et al.: 2004, 'Web Services Notification'. <http://www-106.ibm.com/developerworks/library/specification/ws-notification/>.
- Leymann, F.: 2003, 'Web Services: Distributed Applications without Limits - An Outline'. In: *Procs. Of Database Systems for Business, Technology and Web*.

- Li, Y., H. Ghenniwa, and W. Shen: 2003, 'Integrated Description for Agent-Oriented Web Services in e-Marketplaces'. In: *Proc. of AI2003 Workshop on Business Agents and the Semantic Web*. pp. 11–17.
- McIlraith, S., T. C. Son, and H. Zeng: 2001, 'Semantic Web Services'. *IEEE Intelligent Systems* **16**(2), 46–53.
- Newell, A.: 1982, 'The Knowledge Level'. *Artificial Intelligence* **18**, 87–127.
- OASIS Committee Specification: 2002, 'Business Transaction Protocol, 1.0'.
- Papazoglou, M.: 2001, 'Agent-Oriented technology in support of e-Business'. *Communication of the ACM* **44**(4).
- Papazoglou, M. P.: 2003, 'Web Services and Business Transactions'. *World Wide Web Journal* **6**(3).
- Papazoglou, M. P. and D. Georgakopoulos: 2003, 'Service Oriented Computing'. *Communications of the ACM* **46**(10).
- Papazoglou, M. P. and J. Yang: 2002, 'Design Methodology for Web Services and Business Processes'. In: *Procs. of the 3rd VLDB-TES Workshop*.
- Tuecke, S., K. Czajkowski, I. Foster, J. Frey, S. Graham, and C. Kesselman: 2002, 'Grid Service Specification'. Technical report, Open Grid Service Infrastructure WG, Global Grid Forum. Draft 5.
- UDDI Technical White paper: 2001. <http://www.uddi.org/>.
- Web Service Definition Language. <http://www.w3.org/TR/wsdl/>.
- Wooldridge, M. and N. R. Jennings: 1995, 'Intelligent Agents: Theory and Practice'. *Knowledge Engineering Review* **10**(2).