A System for Distributed Enactment of Composite Web Services

S.J. Woodman\textsuperscript{1}, D.J. Palmer\textsuperscript{1}, S.K. Shrivastava\textsuperscript{1}, S.M. Wheater\textsuperscript{2}

\textsuperscript{1}School of Computing Science, University of Newcastle, Newcastle upon Tyne, UK
(S.J. Woodman, Doug. Palmer, Santosh. Shrivastava)\textsuperscript{1}@ncl.ac.uk
\textsuperscript{2}Arjuna Technologies Limited, Nanotechnology Centre, Newcastle upon Tyne, NE1 7RU, UK Stuart.Wheater@arjuna.com

Abstract. Availability of a wide variety of Web services over the Internet offers opportunities of providing new services built by composing them out of existing ones. Service composition poses a number of challenges. A composite service can be very complex in structure, containing many temporal and data-flow dependencies between their constituent services. However, constituent service operations must be scheduled to run respecting these dependencies, despite the possibility of intervening processor and network failures. The architecture must be scalable, providing a decentralised coordination of service execution rather than based on a centralised scheduler; this is particularly important for services spanning different organisations, where reliance on centralised coordination would be impractical. This paper presents the design and implementation of DECS: a workflow management system for Distributed Enactment of Composite Services. A novel feature of DECS is the separation between specification of service composition and its enactment. A DECS service specification can be deployed either for centralised or decentralised coordination, depending upon inter-organisational requirements. A prototype implementation of DECS has been performed using J2EE middleware. The paper describes the DECS task model for specifying service composition and the middleware services that have been implemented in J2EE for coordinating service execution.

1 Introduction

It is becoming increasingly common for a Web service to make use of Web services offered by different organisations. We term such an inter-organisational service a "Composite Web Services" (CSs). There is much research interest in developing high-level tools for CS creation and management, including service specification languages and run time environments for coordinating the execution of (enactment of) constituent services. This paper presents the design and implementation of DECS: a workflow management system for Distributed Enactment of Composite Services. There are several features DECS that make it novel and distinct from existing workflow management systems for Web service enactment.
1. **Flexible Coordination**: workflow management systems for Web service enactment specify the composition of a CS as a business process. Such a specification should be sufficiently abstract, unchattered with specific details of enactment. Specifying the enactment of the business process is a concrete operation which requires further information that is critically dependent on organisational issues. For example, the organisations involved in a CS may have a peer-to-peer business relationship, in which case, a decentralised enactment seems a natural choice, with each organisation responsible for its part of the process. Whereas in a hierarchic relationship, a centralised enactment may well be deemed more appropriate. DECS supports the separation between the specification of service composition and its enactment, enabling the organisations deploying the service to decide how they wish to coordinate it, rather than the designer of the business process.

DECS provides the option of both centralised coordination as shown in Figure 1 and decentralised coordination, as shown in Figure 2. Should the coordination of the service be spread across multiple servers as in Figure 2 a higher level of fault tolerance is provided. In such cases, each server makes the invocations of the constituent web services for its part of the CS, and communicates via a coordination protocol with its peers to orchestrate the overall execution. Should a coordinating server fail or leave, it is possible to move the CSs which that server was coordinating to another server. The cost of doing so is proportionate to the number of CSs being coordinated and the complexity of those CSs.

![Diagram](image)

**Fig. 1. Centralised Coordination of a Composite Service**

2. **Support for reconfiguration**: It is expected that the execution of a CS could take a long time to complete, of the order of days or weeks, and may contain long periods of inactivity often due to the constituent applications requiring
user inputs. It should be possible therefore to reconfigure a CS dynamically because, for example, services may be moved, machines may fail or users requirements may change. DECS provides basic system support for reconfiguration (see section 2).

3. Support for preserving organisational Autonomy: A further advantage of distributed coordination is that it enables organisations to collaborate whilst maintaining autonomy. When an inter-organisation business process is created, it can be deployed such that each organisation coordinates the parts which they are responsible for and which act upon their internal data. Each organisation is aware of the data relating to those tasks which it is coordinating, and also certain pieces of data which they have requested from the other organisations.

The remainder of this paper is structured as follows: section 2 gives an overview of the system; section 3 describes the task model of DECS; in section 4 we describe the system architecture and its implementation; section 5 we discuss patterns for distributing coordination; section 6 describes related work, finally section 7 concludes the paper.

2 System Overview

We discuss how our system has been designed to meet the application requirements implied above, namely: scalability, dependability, interoperability, dy-
dynamic reconfiguration and flexible task composition. The design of the system has been influenced by our earlier work on the OPENflow distributed workflow management system [1,2].

- Scalability: Once a composite service is deployed, there is no reliance on any centralised service which could limit scalability. The decoupling of the business process definition from the deployment aids scalability too it is possible to reconfigure the business process at run-time if more resources are needed.

- Flexible Task Composition: The system supports a simple yet powerful task model (see section 3). This allows the composition of complex services from simple services located both on the local machine and remote web services. A task can perform application specific input selection (e.g. obtain input from one of several sources) and terminate in one of several outcomes.

- Dependability: Dependability is implemented at two levels in DECS: application level and system level. Due to the flexible task composition model mentioned above it is possible to specify different tasks and processes to handle a variety of exceptional circumstances, for instance, compensating tasks, alternative tasks etc. At the system level, DECS makes use of the facilities provided by modern component middleware (J2EE) to ensure that all information relating to the composite service: its structure, data and state are persistent. All interactions between different modules of DECS make use of transactions to ensure that the data remains consistent and tasks do not interfere with each other. If a coordinating node fails mid process, it is possible to recreate the state of the CS and continue processing from where the failure occurred. This is discussed further in Section 4.6.

- Interoperability: The system has been designed to run in any J2EE compliant application server, thereby supporting system level interoperability. As DECS can invoke web services with arbitrary interfaces, application level interoperability is provided.

- Dynamic reconfiguration: The task model referred to earlier is expressive enough to represent dependencies (dataflow and notification) between tasks. The execution environment exposes low level operations for making changes to the structure of the CS by altering the tasks and the dependencies between them. We are currently working on making this dynamic reconfiguration support transactional to ensure that changes are carried out in a consistent manner despite concurrent reconfigurations and machine failures. This work is being done along the lines of our earlier implementation [2].

3 Task Model

DECS makes use of a flexible task model to describe the structure of Composite Services. The schema for such a task model must be expressive enough to allow arbitrary CSs to be defined. Our schema is defined in terms of tasks, temporal dependencies and data dependencies. A task represents an application specific
A unit of work that requires specified input data and produces specified output data and corresponds to an invocation of a web service. A task instance is modelled as receiving one input message, and sending multiple output messages. The web service invoked is treated as a black box with the input and output data specified by the WSDL document associated with the service [3].

A temporal dependency represents the situation where a down-stream task cannot start until an up-stream task has terminated in a particular state. For example, a goods dispatch task should not be invoked until a payment-capture task has completed successfully. A data dependency indicates that a down-stream task requires input data from the up-stream task. Each data dependency implicitly has a temporal dependency associated with it. For example, a dispatch task requires shipping-address which is part of the output from order task. Implicitly this is not available until the order task has completed [1].

Some of the salient points of the task model which aid flexible composition of CSs are presented below:

- Alternative input sources: Any part of the input data can be acquired from more than one source. This enables the introduction of redundant data sources providing application level fault tolerance. The input data for a task is described at the granularity level of wsd1:part. Support for finer granularity data than wsd1:part is described in Section 7.

- Alternative outputs: A task can terminate in one of several states producing distinct outcomes. In terms of WSDL, one of these states will correspond to an "output" and all others to a "fault". This allows different down-stream tasks to be executed depending on the outcome of an up-stream task. For example, compensation in the event of a fault.

- Compound Tasks: A task can be composed from other tasks. The composed tasks themselves can be simple tasks or compound tasks. This is the principle way of providing abstraction.

- Genesis tasks: A genesis task represents a placeholder for a task structure and is used for on-demand instantiation. This allows the system to only instantiate the parts of a complex task which are necessary, and also allows execution of repetitive tasks.

Each input and output message can contain several parts, each representing a piece of data which can be requested by other tasks. Task t2 in Figure 3 has one input message containing two parts, namely i1 and i2. There are two output messages for task t2, O1 and O2, each containing one part, o1 and o2 respectively. A task begins its life in the wait state, awaiting its input data to be complete. When the input message is complete (i.e., all the data dependency and temporal dependencies are satisfied) the task enters the ready state. If alternative inputs become available simultaneously the one with the highest priority is used. Note that the source of an input part can be from an output message (e.g., d1), or an input message (e.g., d2), the latter represents the case when an input is consumed by more than one task. Temporal dependencies are depicted as a dotted line, for example n1 and data dependencies are shown as solid lines. A compound task
such as \( t_1 \) can contain multiple output messages, with each part having several possible sources. The task will terminate when one output message is complete. If multiple output messages become available simultaneously, the one with the highest priority will be chosen. In section 6 on related work, we compare our task model with the recent proposal for Web service enactment, BPEL [4].

![Diagram](image)

**Fig. 3. A Compound Process**

4 System Design and Implementation

The DECS system has been designed and implemented using J2EE middleware. The current version has been tested to run within the JBoss application server [5].

4.1 Overall structure

The structure of the system is shown in Figure 4. The figure is intended to show how a clients request to execute CS enters the Process Initiator; an instance of the process definition is taken from the Process Definition Repository (PDR) and added to the Process Instance Repository (PIR). The coordinator then uses this data to invoke the web services which compose the CS and inform other coordinating servers of data in which they have registered an interest. Every client request instantiates a new and unique process definition instance. The J2EE technologies used to implement the different modules are shown in Figure 5 which matches the structure of Figure 4.
Using the J2EE environment allows DECS to make use of the rich set of functionality which J2EE application servers provide. For example, uniform access to persistent storage, flexible transaction control, a unified security model. DECS utilises the Entity Beans, Session Beans and Message Driven Beans from the J2EE architecture. An Entity Bean represents a business object that exists in the enterprise application. The application server controls access to Entity Beans, guaranteeing atomicity, consistency, isolation and durability. A Session Bean provides the business logic for a J2EE application. Clients call methods of the Session Bean to interact with the application. Message Driven Beans provide asynchronous behaviour by acting as message listeners for the Java Message Service (JMS). JMS facilitates the exchange of messages among software applications over a network. The prototype of DECS is interoperable in that it will run in any J2EE application server as no proprietary extensions have been exploited.

**Fig. 4. System Architecture**

### 4.2 Process Initiator

The process initiator is able to dynamically deploy web service endpoints which can be used by a client to invoke the web service which corresponds to a process. When a composite service is deployed in the system, an endpoint is generated which allows a client to invoke it. This enables us to achieve transparency from the client’s perspective, whereby the client may be unaware that they are invoking a web service which is implemented as a process. This style differs from many coordination engines which expose one interface to invoke many processes. For example, web service clients of the CARNOT workflow engine must send a message to the WsWorkflowSessionService" endpoint passing the name of the
process to invoke and a context. One of the problems with this style of interaction is that there is no standard way of locating the required service. Processes which can be invoked using the Process Initiator can be advertised using any current advertising method, e.g. UDDI [6].

When designing a CS, one server is designated as the root controller for that CS. This is usually the server which will coordinate the first task in the CS but this is not a requirement. The root controller is the only server on which the Process Initiator deploys an endpoint. When a CS is invoked by a client, an instance of the process definition is created in the PIR based on the process description stored in the PDR. This contains all the data necessary to run the CS, such as the tasks involved and the inter-task dependencies which must be satisfied. The root controller also sends a message to the other servers coordinating the execution requesting that they create an instance of the fragment of the process definition that they are responsible for. This is discussed further in Section 5. The initial values received in the client request are then added to the "root" task in the repository and those tasks whose input dependencies are satisfied are invoked.

4.3 Process Definition Repository (PDR)

The Process Definition Repository stores process definitions and provides a method for instantiating an instance of a process. DECS provides tool support adding an XML based process definition script to the PDR. This can either be a complete process definition for centralised coordination or a partial definition if decentralised coordination is required. The PDR is implemented as a collection of Java Entity Beans deployed in the J2EE application server.

A process definition is represented by a schema which matches the task model described in the previous section, in terms of tasks and dependencies. We have
made use of the concepts of a scripting language developed for a previous distributed workflow system and adapted it to suit web services style invocations. The scripting language was designed to express composite service composition and inter-task dependencies of fault tolerant distributed applications whose executions could span arbitrary large durations [7].

4.4 Process Instance Repository (PIR)

The system will instantiate a process instance based on a process definition for each request from a client. The data related to each process instance is stored in the PIR which is also implemented as a collection of Entity Beans. The state of the process instance is persistent and each change is made persistent, this means that coordination of process instances can be continued after machine failure.

Creating a process instance from a process definition involves three steps:

- Create the local structure of the process instance in the PIR according to the definition stored in the PDR.
- Instantiate the fragments of the schema which reside on remote nodes. This is described further in Section 6.
- Insert the initial values into the process instance from the client request.

4.5 Coordinator

The coordinator in DECS orchestrates the execution of the CS across multiple nodes. This involves checking for input availability, maintaining state of the task and propagating the results both locally and remotely. The prototype uses Session Beans to implement the business logic associated with coordination.

When a process definition is instantiated all the tasks are in the wait state. As input data is added to the tasks they are checked to see if their input message is complete. Once the input message is complete the task moves to a ready state and is put on a persistent JMS queue to indicate that the task is invocable. At this stage, the coordinator checks to see if there are any input dependencies or notifications on the task’s input message. If local dependencies (either data or temporal) exist, the data is propagated locally to these tasks. If a dependency is remote, this initiates a notification of the data to the remote coordinator. These are shown in Figure 8(iii) as d1 and n1 respectively. When propagating data to the remote coordinator, either SOAP or Java RMI can be used. SOAP is intended to be used as the primary communication method but Java RMI can be used to optimise the communications if both coordinators are located on the same network. In both cases, the local coordinator must communicate with the remote coordinator via an RPC style call. The parameters of the notification include the unique identifier of the process instance which is the source of the data and the data value. The action of inserting the data part to the remote task has the side effect of checking the task to see if it can be executed. The semantics of adding any data, local or remote, to a task results in the task being executed as soon as its input message is available. When the output message of
the root task is complete, the Process Initiator will create a SOAP response and send it back to the client.

4.6 Invoker

The invoker is responsible for invoking the web services which comprise the composite service. The invoker is implemented as Message Driven Beans in the application server. This allows us model the asynchronous behaviour that is required to invoke the constituent web services when their dependencies have been fulfilled.

To cope with the consequences of coordinator failure, the designer of the service is currently able to specify one of two actions to be taken on restart for each task which was executing when the failure occurred. These correspond to whether they wish the constituent web services to be invoked at most once or at least once. If the designer specifies they wish to use at most once semantics, transactions are not used in the invoker. The invoker is not aware that an instance of this service was executing when it failed, so it does not attempt to re-invoke the web service request. If at least once semantics are required, everything performed by the invoker is within one transaction. This includes obtaining the input data for the web service, invoking that service, receiving the results and adding them to the PIR. If the invoker fails at any point within this transaction, it will be rolled back on restart. This results in the web service being invoked with the same input data again, once the system has recovered. If an end-to-end transaction protocol were available, it would be possible to achieve exactly once execution semantics for the constituent web services of the process. This is because any failure in the coordinator would be propagated to the constituent web services which could rollback their execution too. On restart, the invoker would send the SOAP request again and eventually a successful execution would occur.

5 Distribution Patterns

It is the responsibility of the designer of the service to specify how they wish to distribute the coordination of the service. Simple CSs may be coordinated centrally, as shown in Figure 6. This is the simplest scenario, where a central node performs all the invocations of the services necessary to complete the CS. More complex scenarios can be built, where the coordination of the CS is divided between multiple nodes which run DECS. One example is given in Figure 7 where the service is divided between two nodes. How to divide the service is an application specific decision. If we consider the notion of a Virtual Enterprise where a set of companies wish to collaborate to provide a service, the division of coordination could be along organisational boundaries. This allows each organisation to coordinate their own part of the service, and possibly utilise their private services. For example, company S wishes to sell a product and use company D to deliver the product. The composite service could be divided such that nodes at S coordinate the ordering of the product and payment and then the nodes at D
coordinate scheduling of the delivery of the product. A division such as this has some advantages: firstly, S can integrate the order with their own procurement process allowing re-ordering of stock if necessary and D can integrate with their private delivery scheduling services; secondly, data security is higher as only the minimum amount of data is passed across from one organisation to the other to allow the CS to continue its execution.

![Diagram of Centralised Coordination](image)

**Fig. 6. Centralised Coordination**

The system aims to give each node the minimum information necessary to coordinate the execution of that part of the CS. Each node only stores the data about the tasks which it is coordinating, the internal dependencies which must be satisfied and the external notifications which must be sent and received. The node is not aware of what the other nodes are doing, or what tasks they are coordinating. This is intended to provide autonomy; it makes it attractive to businesses that do not wish to disclose all of their internals but do wish to integrate their business processes with a trading partner.

Figure 8 shows an example of how a very simplistic CS, A could be divided to run over three servers, X, Y and Z. Figure 8(i) shows the overall CS with figures 8(ii) to 8(iv) showing the three servers views of the service. Server X is delegated as the controller of the CS so it is this server which exposes an endpoint allowing the service to be invoked. When a client request is received at X for service A, the first task to be executed will be B. When B completes, there are no dependent tasks at server X, but there is a notification request for the results to be sent to Y. When Y receives the results of B via a notification, task C is executed (by Y) and on completion the results sent via a notification to Z. As there is also a dependency for task E at Y, the data is propagated locally to task E’s input message. As the input message for task E is not yet complete, the task is not invoker. At server Z, task D’s input message is complete by receiving
the notification from Y so task D is executed. A notification is sent from Z to Y on completion of D which causes the input message of task E to be complete and thus task E fired. Completion of task E causes a notification to be sent to X with the results which are used as input for task F. On completion of task F, the output message for the compound task A is complete, so the result is sent to the client.

6 Related Work

There are several industry led efforts aimed at developing (often competing!) standards for specifying, composing and coordinating the execution of CSs. The specifications are still evolving and often ill defined [8]. We compare our task model with a recently proposed Web service execution language standard. The aim of Business Process Execution Language for Web Services (BPEL) [4] is to provide a standard for specifying business process behaviour and business process interactions, for applications composed from Web services. We are going to focus on a comparison of the approach taken by BPEL for specifying business process behaviour and that taken by DECS.

Like DECS the BPEL process model allows business partners interact through peer-level conversations, using both synchronous and asynchronous messages. These conversations are carried out between the partners using specified sets of Web services. BPEL has been designed to allow the coordination of distributed web services in a centralised manner. The coordination itself was not designed to be distributed. The coordination model is tailored towards implementations that have a centralised state upon which a rich set of activities can act. The result is that providing a decentralised enactment engine is difficult especially if
the intention is to deploy the workflow in a wide area network. In comparison, the DECS task model was designed such that each task holds a small amount of state and there is a minimal set of services which are able to act on this state. This results in a model which is easier to distribute as each task encapsulates its own state and it is easier to migrate task coordination to another node.

BPEL relies on specifying a large set of explicit control flow activities such as forking, joining and conditionals. Conversely, control flow in DECS is implicit and specified through dependencies. These constructs have been found to be sufficient to specify complex processes. Another consequence of BPEL having such an extensive set of features which can be combined to specify a process is that the resulting process cannot be easily analyzed to verify such properties as eventual termination. The DECS model is much simpler and amenable to analysis, allowing both temporal and correctness properties to be checked [9].

There are a number of systems available which coordinate the orchestration of composite services produced both in the academic community and in industry. Some of these are discussed below and briefly compared and contrasted to DECS.

Collaxa [10] is well-known product that can enact composite services specified in BPEL. It is a centralised coordination engine; this is inevitable, given the observations on BPEL above. As such, it cannot offer the same level of flexibility in deployment scenarios, dependability and scalability that DECS intends to provide.
Both eFlow [11] and BioOpera [12] explore the declarative composition of services but concentrate on a centralised orchestration model. Successful efforts have been made in eFlow to allow dynamic refactoring of services although this is simplified due to a central, global view being available.

DECS has been designed to run in a J2EE environment to allow portability and ease of integration into existing enterprise applications. DySco [13] offers many similar features to DECS, such as decentralised coordination but portability has not been addressed to the same level; it is not designed to run in standard middleware such as J2EE. Conversely, CARNOT [14] offers portability through a J2EE implementation, but does not support distributed orchestration of composite services.

Another approach which has been explored is that of SELF-SERV [15], based on a declarative state-chart oriented language. It also includes a peer-to-peer coordination model and provides support for equivalent services through the use of service communities. It comes closest to DECS in terms of design aims and functionality.

7 Concluding Remarks

We have presented the design and implementation of DECS: a workflow management system for Distributed Enactment of Composite Services. A novel feature of DECS is the separation between specification of service composition and its enactment. A DECS service specification can be deployed either for centralised or decentralised coordination, depending upon inter-organisational requirements. A prototype implementation of DECS has been performed using J2EE middleware.

A suite of common services is being developed as part of the DECS. Such services include:

- User input service: it is likely that some CSs will require input from users at different parts of the execution. For this reason we are developing a servlet based user interface which will allow users to input parameters to be used in the execution. The data entered may be used to determine the consequent flow of execution or to provide advanced error recovery.

- Send and Receive services: In order to allow asynchronous communications services will be developed which will send or receive a message. Tasks which utilise these services can be added to any CS, thus potentially providing a fully asynchronous CS. We envisage more web services becoming available which require message based communications rather than the RPC style services which are common at present. Such document exchange web services are more versatile and allow easier integration of business processes. However, with the introduction of this style of interaction problems are introduced such as message correlation and temporal issues. These will be investigated further.

- Administrative Services: services will be provided which allow a user (with appropriate permissions) to deploy, remove and dynamically reconfigure pro-
cess definitions. Care must be taken when refactoring a service which is distributed across multiple nodes to ensure that deadlock is prevented. This could occur in cases where tasks are removed upon which a remote task is awaiting a notification. The service will provide mechanisms to ensure that this situation does not occur.

- Transformation of complex types: At present, the system is able to manipulate the flow of data at the granularity of WSDL parts. However, if a complex type is defined in WSDL the system treats this as a black box and cannot address internal fields. We see this as a deficiency in the system so aim to provide a service which is able to transform complex types so that they conform to another schema. This is likely to be done using XSLT, with the designer of the CS providing a style sheet describing the transformation required. An example where this would be useful would be extracting the invoice-number from an invoice type that was returned from the order service and use it as the input to another service.

Acknowledgements

Discussions with Gustavo Alonso clarified our ideas. This work is part-funded by the UK EPSRC under grant GR/N35953/01: Information Co-ordination and Sharing in Virtual Environments; by the European Union under Project IST-2001-37126: ADAPT (Middleware Technologies for Adaptive and Composable Distributed Components); and by the UK DTI e-Science programme under project “GridMist: Middleware Services and Tools for Managing Resource Sharing in Virtual Organisations”. This work will be continued in the EPSRC funded project GR/S63199: “Trusted Coordination in Dynamic Virtual Organisations”.

References