A Workflow System through Cooperating Agents for Control and Document Flow over the Internet *

A. Dogac[1], Y. Tambag[1], A. Tumer[1], M. Ezbiderli[1], N. Tatbul[1], N. Hamali[1], C. Icdem[1] and C. Beeri[2]

[1] Software Research and Development Center
Middle East Technical University (METU), 06531 Ankara Turkiye
asuman@srdc.metu.edu.tr
[2] Hebrew University, Jerusalem, Israel
beeri@cs.huji.ac.il

Abstract. In this paper we describe an architecture that provides for automating and monitoring the flow of control and document over the Internet among different organizations, thereby creating a platform necessary to describe higher order processes involving several organizations and companies. The higher order process is designed through a graphical user interface and is executed through cooperating agents that are automatically initialized at each site that the process executes. Agents handle the activities at their site, provide for coordination with other agents in the system by routing the documents in electronic form according to the process description. The system is capable of activating external applications (which may be inside the company firewall) when necessary, keeping track of process information, and providing for the security and authentication of documents as well as comprehensive monitoring facilities. The architecture is general enough to be applied to any business practice where data flow and invocation of activities among different industries and cooperations follow a pattern that can be described through a process definition, however since the project is on maritime industry, some of the graphical user interfaces are customized accordingly. The system is fully operational for industrial use.

1 Introduction

In the MARIFlow system described in this paper, the higher order process is defined through a graphical user interface which is then mapped to a textual language called FlowDL. FlowDL is a block structured language encapsulating the six primitives defined by the Workflow Management Coalition through its blocks with which it is possible to describe flows and hence construct a workflow

* This work is being supported by the European Commission Project Number: INCO-DC 97-2496 MARIFlow, by the Middle East Technical University Project Number: AFP-97-07-02-08, and by the Scientific and Technical Research Council of Turkey, Project Number:197E038

specification [3]. FlowDL allows to indicate the source of the documents, their control flow and the activities that make use of these documents.

A MARIFlow process is executed through cooperating agents, called MARCAs (MARIFlow Cooperating Agents) that are automatically initialized at each site that the process executes. MARCAs handle the activities at their site, provide for coordination with other MARCAs in the system by routing the documents in electronic form according to the process description, keeping track of process information, and providing for the security and authentication of documents as well as comprehensive monitoring facilities.

The responsibilities of the agents (MARCAs) in our architecture are as follows:

- A MARCA receives messages through a persistent queue and evaluates them to decide what specific action to take.
- It persistently stores the documents it receives. It should be noted that the organizations may be reluctant to grant access inside the corporate firewall. In such cases when the need arises, the MARCA passes these documents to an in-house system by properly acknowledging the in-house system on further processing that may be necessary on the documents. The MARCA is also responsible for getting the documents from the in-house system and forwarding them to the related agents as specified in the process definition.
- Process related information also needs to be stored persistently for monitoring purposes. In our system MARCAs store the information related with monitoring to any JDBC compliant database to be accessed through a JDBC interface.
- There is a single MARCA at each site that handles all the activities of all workflow definitions and their instances related with that site. Therefore a new MARCA is generated only for a site participating to the system for the first time.

If we summarize, the functionality provided by the system developed is as follows:

- A declarative means to specify the control and document flow over the Internet where it is possible to define the source of data, its control flow and the activities that make use of this data.
- Invoking external applications (which may be inside the company domain) when necessary.
- Authentication and security of documents and the process related information.
- A monitoring mechanism for keeping track of the documents and/or for providing detailed account of the current status of a process instance within the system.
- Measures for failure recovery and exception handling.

In the following the system is described very briefly due to space limitations. Interested reader is referred to [2] where full design and implementation details are presented.

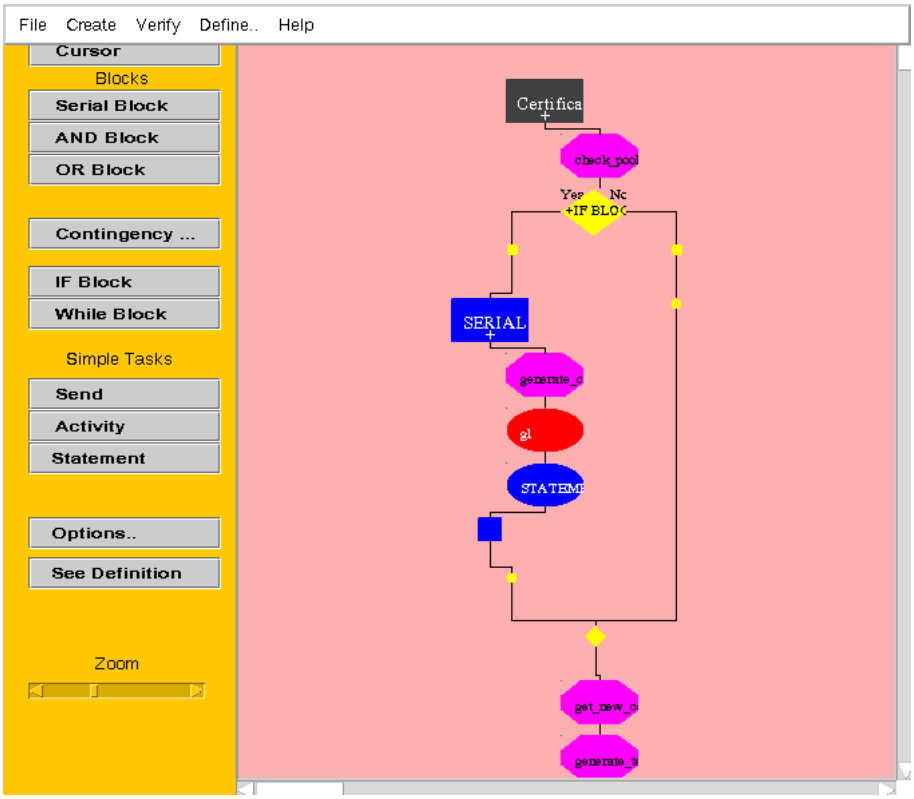

Fig. 1. The Graphical User Interface for Process Definition in MARIFlow

2 The Architecture of the System

In MARIFlow system each organization may have in-house applications inside a firewall protected from unauthorized accesses. MARCAs exist outside the firewall and inform in house applications when necessary through a User MARCA Interfacing Application (UMIA).

An inter enterprise workflow is defined graphically as shown in Figure 1. This tool allows the workflow designer to specify domains, tasks and process information which are then used in building the process definition graphically. This definition is mapped to the textual FlowDL language. The information on sites at which a MARCA should be installed are obtained from the domain definition in the process specification. These sites download a generic MARCA from a given URL. At compilation time the guards of activities within the responsibility of a MARCA are also determined according to the process definition and the MARCAs are initialized with these guards through the Workflow Definition Tool.

Guards are logical expressions for significant events of activities of a MARCA like "start" and "terminate". MARCAs evaluate these guards with the messages that they receive to decide on their actions. As an example, the start guard of an activity handled by a MARCA can be the arrival of a document, say, "doc1" from site "S1" and a document "doc2" from site "S2". In this case, MARCA will start execution of this activity when this AND expression evaluates to true by the arrival of the mentioned documents. Clearly the guards are generated from the information given in the process specification. Similarly "terminate" guard of an activity handled by MARCA may require the transmission of a document, say, obtained from the in-house system to another MARCA. It should be noted that this transmission is realized through persistent queues to survive through crashes.

3 MARIFlow Security Services

Within the life cycle of a process definition an application in the company domain may prepare a message, then may pass it to local UMIA. UMIA, in turn, passes it to the corresponding MARCA. The MARCA sends the message to one or more other MARCAs according to the process definition. The receiver MARCAs pass the messages to their corresponding UMIAs, which pass them to the local applications. Thus, inter/intra company communications basically consists of communication sessions between MARCAs and between a MARCA and its UMIA.

The security requirements of the system can therefore be summarized as follows:

- Confidentiality of data transfer between MARCAs: The contents of a message, including the process description fields, should not be visible to anybody except the sender and the receiver.
- Confidentiality of data transfer between a MARCA and the corresponding UMIA, as above.
- Authentication and integrity for both levels.
- Signatures for some of the inter/intra-company messages, such as certain types of test and certification documents. Here, a message consists of a document that is passed in order to be stored.

A comprehensive set of security services has been developed and integrated into the system which are described in [2].

4 Failure and Exception Handling

MARIFlow system provides comprehensive set of measures for failure recovery and exception handling.

4.1 Recovery of MARCAs

When a site goes down, restarting the MARCA is under the responsibility of the Operating System's start up control. The site's start up control analyzes the persistent logs of the MARCA and start a new instance using these stable logs created before the site crash. However, there is need for a further mechanism to prevent any Operating System related problem.

In Mariflow, for each MARCA installed there is a background process at that site, called the "rescue process". The rescue process is responsible for monitoring the life time of the agent and checks the MARCA at specific time intervals through a predetermined socket. A thread of the MARCA listens to this socket and responds to the signals. If the MARCA does not respond to this process for a given period of time, the process starts sending signals more frequently. If the MARCA still does not respond, after sending a bunch of signals the process assumes that the MARCA is not functional. The two possibilities in this case are: the MARCA could be blocked or it could be dead. When the rescue process is unable to find the OS process that belongs to this MARCA (i.e., it is dead), it instantiates a new MARCA by the help of the persistent logs related with the state of the agent.

Otherwise if the MARCA is blocked, it is necessary to kill the old instance prior to installation of a new instance. Since the logs are persistent it is possible to recover the state of the MARCA killed and hence the site does not suffer from any inconsistencies.

For the described mechanism to work correctly it is necessary to make sure that rescue process stays alive. Therefore, just as the rescue process checks to see that the MARCA stays alive, the MARCA also checks to ensure that the rescue process stays alive by signalling the rescue process at predefined time intervals. It is MARCA who reinstantiates the rescue process when it dies.

4.2 Failure Handling

The hierarchical approach to failure handling described in [1] is implemented in MARIFlow system which allows for partially rolling back the workflow instance to the nearest point in process history tree where it is possible to restart the execution. When a sub-activity T fails, it is necessary to determine the impact of that failure on the ancestors of T by finding out the highest level ancestor that should be aborted. The details of this technique is given in [1].

4.3 Exception Handling

The MARIFlow system handles the following types of exceptions:

- Semantic exceptions occur when a deviation from the expected behaviour in the program logic is encountered. These are handled through the IF block.

- Exceptions Raised by the Communication Infrastructure: Various types of errors can be encountered during communication between two agents, or communication with a database system or a mail server. The communication system recovers from possible failures by retrying the operation when possible and informing the user program if the request cannot be issued.
- Exceptions caused by NON-VITAL Activities: It should be noted that when document flow is a part of a workflow system, more often than not, there will be a need to archive the documents. The transfer and archival of the documents may take considerable amount of time. Therefore a mechanism which allows the other activities in the system (that does not use these documents) to proceed without waiting these archival activities provides for better performance. Yet there should also be mechanisms to guarantee that the workflow instance will not terminate before the successful termination of all such activities. We use NON-VITAL activities suggested in [1] with some modification. Originally the NON-VITAL activities are defined to be those, whose failure does not effect the flow of the process. On the other hand we say that NON-VITAL activities are those that can be assumed to terminate as soon as they start but raise an exception when they fail. In this way a NON-VITAL activity does not delay the execution of other activities unnecessarily; yet their successful termination is guaranteed by the exception handling mechanism.

5 Conclusions

MARIFlow system implements a fully distributed inter-enterprise workflow system through cooperating agents. The system is developed within the scope of the European Commission supported INCO-DC 97 2496 MARIFlow project and is fully operational for industrial use.

References

1. Q. Chen, U. Dayal, "A Transactional Nested Process Management System", in Proc. 12th International Conference on Data Engineering, New Orleans, 1996.
2. A. Dogac, Y. Tambag, A. Tumer, M. Ezbiderli, N. Tatbul, N. Hamali, C. Icdem and C. Beeri, "A Workflow System through Cooperating Agents for Control and Document Flow over the Internet", Technical Report, Software Research and Development Center, METU, March 2000, http://www.srdc.metu.edu.tr/publications.html.
3. D. Hollingsworth, *"The Workflow Reference Model"*, Technical Report TC00-1003, Workflow Management Coalition, December 1996.

This article was processed using the LaTeX macro package with LLNCS style