A Petri Net-based Model for Knowledge-based Workflows in Distributed Cancer Therapy

R. Müller
Institut für Informatik, Universität Leipzig
mueller@informatik.uni-leipzig.de

B. Heller
Institut für Medizinische Informatik, Statistik und Epidemiologie, Universität Leipzig
heller@imise.uni-leipzig.de

In: Proceedings of the International EDBT98 workshop on Workflow Management Systems
Valencia, Spain, March 1998

Abstract

This paper describes a Petri Net-based model for workflow management in distributed cancer therapy, a domain which is characterized by the involvement of several geographically distributed departments and a long-term treatment. The approach uses hierarchical Petri-Nets; tokens, places and transitions are objects conform to the CORBA/IDL-model to facilitate implementation in a heterogeneous environment. During treatment, generic organizational nets are dynamically refined with disease-specific subnets obtained from an oncological knowledge base to adapt the net in a patient-specific way. This refinement task is mainly processed by a *Knowledge Refinement Manager* which has been added to a commercial workflow architecture. Currently, a prototype implementing the model, called HEMATOWORK, is developed at Leipzig University.

1 Introduction

A major problem impeding the integration of medical information systems and expert applications into clinical environments is the lack of support concerning the associated organizational and medical workflows (e.g. [1][2]). For example, the knowledge-based system THEMPO [3] provides a detailed chemotherapy calculation and diagnostic monitoring for cancer patients, but does not support related processes such as making diagnostic appointments in other departments, sending test material to the central laboratory or external specialists, or managing the medical reports that have to be distributed to several sites within and outside the hospital. Workflow management, with its explicit representation of long-term processes and integrating view of automated and manual activities (e.g. [4][5]) has been identified as a promising approach to overcome these limitations not only in medical domains.

However, workflow technology so far has neglected the problem of domain knowledge influencing working processes, or implicitly assumes that knowledge about how to deal with cases (e.g. patients) is represented by the process definitions themselves or encoded within the involved applications. This assumption may hold in simple domains, but is inadequate in the "knowledge-intensive" domain of medicine, where nearly every patient-oriented process is influenced by knowledge about the particular disease of this patient. For example, the number and types of diagnostic procedures checking whether a patient will tolerate an aggressive chemotherapy significantly depends on the particular disease of the patient and the selected chemotherapy. Therefore, a clear separation of process definitions on the one side and disease-specific knowledge within a knowledge base on the other side must be viewed as essential for workflow management in medical domains, because of the following basic reasons:

- Representing all medical possibilities within a workflow definition would lead to an exorbitant number of case distinctions and branching nodes within the workflow specification, mixing up organizational processes and disease-specific domain knowledge, and hampering workflow readability and maintenance.
- Secondly, and more important, domain knowledge about diagnostic and therapeutic procedures is also required not only by the workflow management system, but also by other applications using different problem-solving methods. For example, in the domain of hemato-oncology, knowledge about the appropriate sequences of chemotherapy cycles for leukemia patients is required by monitoring tools too, which are used by external expert panels to retrospectively check whether a particular patient has been treated according to the medical standard. This "multiple" usage of domain knowledge is best supported by encapsulating it in a separated

knowledge base using a generic representation mechanism making as less as possible assumptions about the specific execution model of the application requesting the knowledge.

This article reports ongoing work done within the project *Knowledge-based Workflow Management for Distributed Hemato-Oncology* carried out by the University of Leipzig and several University Hospitals. The overall aim of the project is to develop a distributed knowledge-based workflow system - called HEMATOWORK - which supports the medical staff with respect to the treatment of hematological tumors and the management of the associated processes, including data-intensive communication and material transfer between local and external departments and specialists. The term "knowledge-based" indicates, that - at runtime- HEMATOWORK communicates with a knowledge base via a *Knowledge Refinement Manager* to episodically refine and adapt (parts of) a workflow template with respect to the patient's specific situation.

The paper is organized as follows: Section 2 briefly discusses related workflow projects in medical domains, while the domain of distributed hemato-oncology is characterized in section 3. Section 4 describes the modeling approach, which is based on object-oriented hierarchical Petri Nets. This net language is then used in section 5 to model the treatment procedures and data management processes in the domain of hemato-oncology. Section 6 discusses architectural and implementation issues, including the architecture of a PN-based Workflow Management System to implement the model. Section 7 concludes the paper with a summary and open problems of the approach.

2 Related Work

Workflow technology already has been applied to several medical domains to support clinical processes and distributed medical data processing. Sheth et al. [6], for example, have developed a large-scale distributed workflow system, based on CORBA/Web technology, to support the state-wide immunization process and tracking in the State of Connecticut. Lee et al. [7] have implemented a hospital care planning system based on the OZ Collaborative Environment. Both approaches, however, do not provide knowledge modeling techniques sufficient for the complex domain of oncology. Dazzi et al. [8] introduce an approach characterized by a clear separation of organizational and medical processes for leukemia treatment, but abstract of aspects of heterogeneity and distribution which are essential in the scenario of distributed hemato-oncology described below.

3 Medical Background and Domain Analysis

Modern hemato-oncology is characterized by the usage of standardized *protocols* containing diagnostic criteria and detailed instructions about chemo- and radiotherapy. Generally, for every relevant cancer type (lymphoma, leukemia etc.), there exists a specific protocol which is compiled and edited by an expert commission and then distributed to the treating hospitals. Furthermore, the treating institutions send diagnostic and therapeutic reports about the patients back to the central commission. These reports are then collected and statistically analyzed to identify for example if new drug combinations have increased the survival rate significantly. Additionally, diagnostic material and therapeutical plans are sent to external expert panels (e.g. pathologists) supporting the local physicians.
However, the introduction of protocol-directed care has gone hand in hand with an increasing complexity of the diagnostic and treatment processes, and has led to new challenges and burdens for the medical staff, especially with regard to medical decision making and data management. Therefore, a distributed workflow system, having explicit knowledge about the treatment and the associated communication paths between the involved institutions, is assumed to provide useful support for the medical staff. In particular, the workflow system HEMATOWORK intends to support the following basic tasks:

- ***Treatment Functionality:*** The support of diagnosis and long-term treatment of hematological cancer patients at the oncological ward or ambulance is viewed as the core service of HEMATOWORK. This functionality for example includes chemotherapy and radiology management as well as diagnostic monitoring during the aggressive therapy, and is achieved through a suite of specialized applications (e.g. for calculating chemotherapy dosages) and databases controlled and coordinated by the system.
- ***Intra-hospital Communication Functionality:*** As every specialized medical workflow system inherently requires services of other local sections and departments, HEMATOWORK provides processes for communication and material transfer between the oncological site and other local departments via the Hospital Information System. This includes, for example, appointment management with departments such as the radiological section, sending diagnostic test material to the histological department or the hospital laboratory, or ordering drugs from the local pharmacy.
- ***Inter-hospital Communication Functionality:*** This covers the communication and material transfer between the treating hospital and external expert panels and central commissions. It includes sending diagnostic and treatment reports and test material from the hospital to external

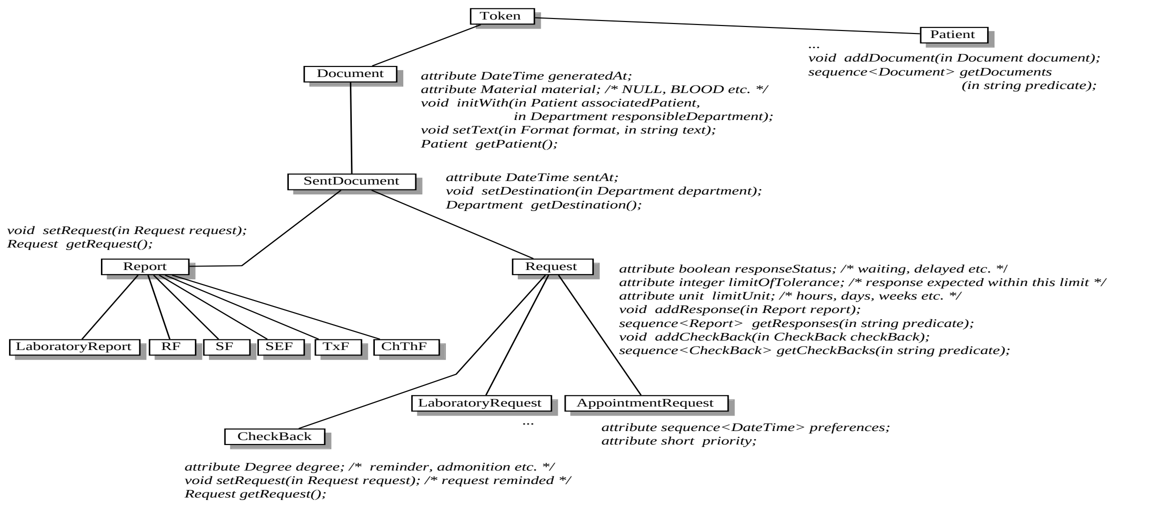


Figure 1: Sample token class hierarchy (including parts of the associated IDL-interfaces) for the medical documents managed by the workflow system. *Request* and *Report* form a 1:n-relationship, a *Request* can be reminded by a *CheckBack*. RF, SF, SEF, TxF and ChThF are special hematological reports generated and processed during treatment (RF=Oncological Registration Form, SF=Staging Form, SEF=Supplemental Examination From,TxF=Toxicity Form, ChThF=Chemotherapy Form). See figure 2 for a PN processing document tokens of this class hierarchy.

specialists and commissions. This functionality is called *inter*-hospital, as these commissions and specialists usually are located at hospitals in other towns or states in Germany.

- ***Tracking Functionality:*** This part of HEMATOWORK is located at the central commissions, and manages the tracking of the patients treated by the hospitals. This involves periodical checks whether all reports have been received for a particular patient, and includes generating reminders and admonitions if reports are missing. Furthermore, incoming data are checked with respect to medical plausibility and correctness.

4 Object-Oriented, Hierarchical Petri Nets

Petri Nets (PNs) have been extensively used to formally describe, analyze and simulate the behavior of dynamic systems and concurrent processes. As the clear and formal description of the supported processes is a major requirement for reliable workflow applications, PNs and especially extensions such as Colored and Hierarchical PNs have been used as a formal foundation for workflow models [9] and workflow management systems [10][11].

This section briefly introduces the PN constructs that have been used in this article to model the workflows and communication paths in the domain of cancer therapy. The PN language described is a combination and extension of hierarchical and object-oriented PNs based on [12][13]. In particular, beside the “standard” constructs such as transitions, places, firing conditions etc., the following extensions are used:

- To allow flexible data modeling and to facilitate an implementation of the PN workflow model within a heterogeneous and distributed environment, tokens are *objects* conform to the OMG/CORBA-model, being organized in class hierarchies and providing IDL interfaces with public operators used by the transitions. Figure 1 gives a sample token class hierarchy, and figure 2 shows a PN processing token instances of the classes *Request* and *Report* (or their subclasses) of figure 1.
- Places and transitions can be subclassed; instances of these classes may be - for example - constrained concerning their token input and output and usually induce a specific behavior of the net interpreter. An important transition subclass of the oncological model is, for example, *Determine-AndExecuteMedicalPlan*. Instances of this class, which are placeholders for diagnostic and therapeutic plans, induce a request to the knowledge base to determine an appropriate medical plan for a patient, and are then refined with the retrieved plan (see section 5 for further details). An example for a place subclass is *InterHospitalComm* (for Inter Hospital Communication); instances of these places represent communication points between (geographically) separated departments, and are usually implemented by a Web-based communica-

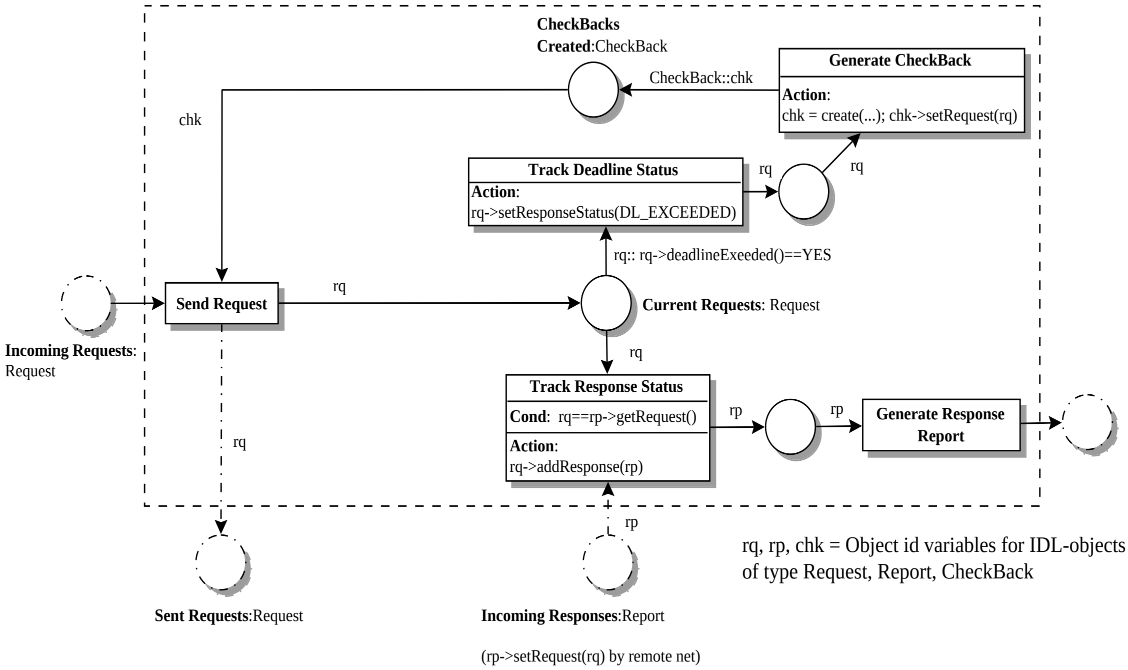


Figure 2: Simple PN example: This net sends requests, keeps track whether there is a response within an expected time interval, and - if not - generates check-backs (and - if necessary - check-backs on check-backs or admonitions) to obtain the requested information or material. This net is part of nearly all nets managing the medical and non-medical communication between the departments (oncological ward and ambulance, central laboratory, histological department, central commission and external expert panels) involved in cancer therapy. The string after the colon of a place label denotes the token class(es) allowed in a place. The labels of the edges denote object references/constraints used by the transitions (with the format "[Class::]object-reference::condition").

tion. Similar to tokens, places and transitions have a CORBA/IDL-Interface.

5 A Petri-Net Workflow Model for Distributed Cancer Therapy

5.1 Organizational Processes

First of all, HEMATOWORK provides a set of generic workflow nets expressing the organizational structure of distributed hemato-oncology and abstracting from the particular oncological diseases and therapeutic procedures. Processes of this organizational class mainly cover disease-independent aspects such as patient admission and discharge, or document management and material transfer between the involved departments. Figure 3 for example shows the global communication paths and abstract high-level transitions of distributed hemato-oncology.
As it is expected that additional hospitals and specialist practices will participate in the oncological protocol-directed trials in the future, this organizational model does not make any assumptions about the number or geographic location of the participating organizations (*organizational* scalability). For example, the organizations and departments illustrated in figure 3 may be located at the same site, or in different towns. To facilitate a site-specific implementing, the model specifies all data requirements in CORBA/IDL.

5.2 Disease-Specific Processes and Knowledge

Disease-specific aspects are linked to an organizational net by cancer-type specific token classes (such as leukemia specific *Report* token subclasses derived from the generic *Report* class in figure 1), or by refining the net during run-time through disease-specific nets obtained from the oncological knowledge base. To separate organizational from disease-specific processes and to support multiple usage of medical knowledge (as knowledge is used not only by the workflow system, but also by other applications), disease-dependent knowledge is encapsulated in a knowledge base. This base contains the following layers of non-procedural

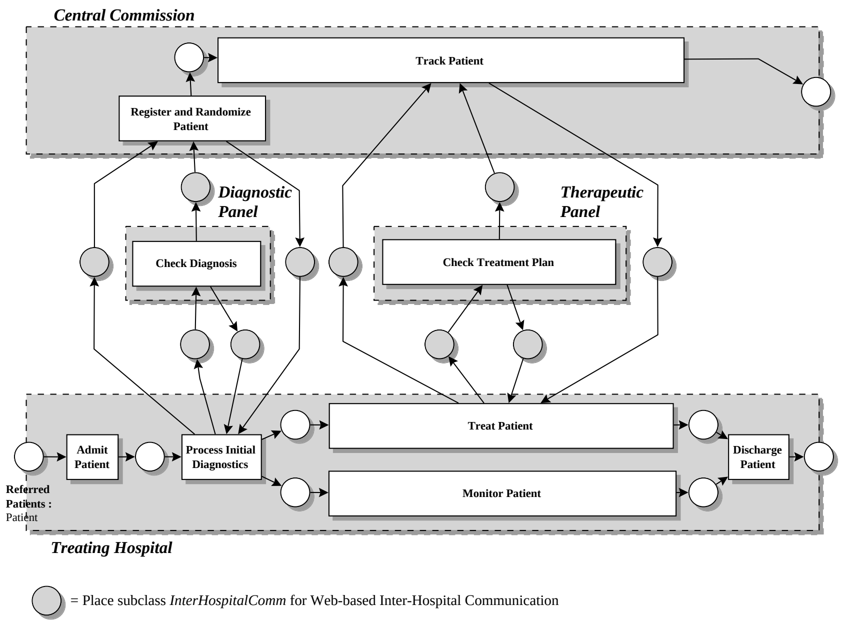


Figure 3: Top-level PN describing the overall global communication paths between the involved departments. All transitions are abstract in the sense that they are refined trough large subnets. The gray places are instances of the class *InterHospitalComm* (for Inter Hospital Communication), and are usually implemented as Web-based communication points between geographically distributed departments. The *Treat Patient* transition is refined by a subnet containing abstract transition expressing the different phases of therapy (e.g. *Initial Therapy, Consolidation Therapy* etc.), which are common for all hematological cancer types (see figure 5 for the knowledge-based refinement of an *Initial Therapy* transition).

and procedural medical knowledge (see also figure 4):

1. Terminological knowledge as the basis for all other types of knowledge (including the formal definitions of basic medical terms such as *leukocytes* and *leukemia* and their relationships).
2. Knowledge about the basic diagnostic and therapeutic procedures, such as *The chemotherapy CHOP21 consists of the drugs Cyclophosphamid, Vincristine etc. with the daily dosages of 750 mg/m^2, 2 mg etc.*
3. Procedural knowledge about the disease-specific conditional sequences of diagnostic and therapeutic procedures (e.g., lymphoma patients have to be treated with 6 cycles of the chemotherapy CHOP21, and, if tumor is “bulky”, with additional radiation therapy).

For knowledge representation, a semantic network approach has been chosen. For all knowledge layers, there exist acquisition tools which are used by the physician themselves to define the basic medical steps and to assemble larger diagnostic and therapeutic procedures.
Especially the knowledge of layer 3 is dynamically linked to a workflow at run-time for patient-specific workflow adaption. Although the procedural knowledge of this layer is ”close to the workflow view” and principally also could be stored in the workflow library of the management system, it is located in a knowledge base because of the following reasons:

- The definition of control structures of medical procedures in layer 3 has to reference objects of the two layers below. Therefore, the acquisition and maintenance process is facilitated by locating all layers in one base.
- As already mentioned, especially expert system applications (which are controlled by the workflow system) frequently need information from the knowledge base too (including layer 3), but have a totally different problem-solving behavior than the workflow engine. Therefore, the “run-time-oriented” PN notation[1] would not meet their requirements, with the consequence that a more generic representation of medical procedures making as less as possible assumption about the particular execution model is more appropriate in a “multi-application” scenario.

In the following, the disease-specific net refinement process is described in detail:

1. Especially places, with their basic functionality as token repositories, represent a run-time oriented structure, which is not necessary with respect to pure semantic net knowledge representation approaches.

As an oncological treatment usually lasts at least 6 months up to two years, and depends on the specific tumor subtype and the individual therapy response of the patient, the particular sequence of diagnostic and therapeutic procedures can not be known in detail when a patient is admitted. Therefore, the high-level PNs processed by the workflow engine contain several abstract transitions of the type *Determine-AndExecuteMedicalPlan*. The predefined **Action**-part of transitions of this type consists of a *determineMedicalPlan*(..) invocation, which takes a token object of type *Patient* as input parameter and returns a PN-representation of an appropriate medical plan, and a *refine(...)* invocation which refines the *DetermineAndExecuteMedicalPlan* transition with the PN returned by *determineMedicalPlan*(..) (both operators are components of the IDL-interface of this transition class).

When calling *determineMedicalPlan(...)*, control is shifted from the workflow engine to the so-called *Knowledge Refinement Manager* (KRM), which is an agent mediating between the workflow layer and the knowledge base. First of all, the KRM identifies the oncological situation of the patient represented by the *Patient* token (by requesting, for example, the diagnosis and cancer subtype, the sequence of medical procedures already applied to the patient, and known contra-indications). After that, by exploring the oncological semantic networks of the knowledge base, the KRM determines the appropriate subnet representing the therapy steps usually applied to patients with the identified characteristics. A translator, which is a submodule of the KRM, then generates a PN-representation of this subnet retrieved from the knowledge base (by adding, for example, place constructs to the semantic net). The last step then consists of refining the *Determine-AndExecuteMedicalPlan* transition with the retrieved PN, which is then processed by the workflow engine. PNs expressing patient-specific medical procedures may again contain *DetermineAndExecuteMedicalPlan* transitions, which induce the same mechanism again on a finer level of granularity. In figure 5, for example, the PN used for the refinement of a lymphoma therapy contains transitions of type *Determine-AndExecuteChemotherapy* (a subclass of *Determine-AndExecuteMedicalPlan).* When a transition of this type is processed, the KRM retrieves the detailed substructure of the appropriate chemotherapy from the knowledge base, which is again integrated into the PN. By this episodical and recursive mechanism, the hierarchical structure of cancer therapy, which is also reflected by the several layers of the knowledge base, is supported in a natural manner.

6 Architectural and Implementation Issues

Currently, a prototypical implementation of the workflow system HEMATOWORK is realized at Leipzig University, based on IONA ORBIX and COSA WORKFLOW, a PN-based workflow management system [10]. To achieve independence and autonomy of the involved departments, the net in figure 3 has been split up into 4 parts (indicated by the gray rectangles), each part having its own workflow server and

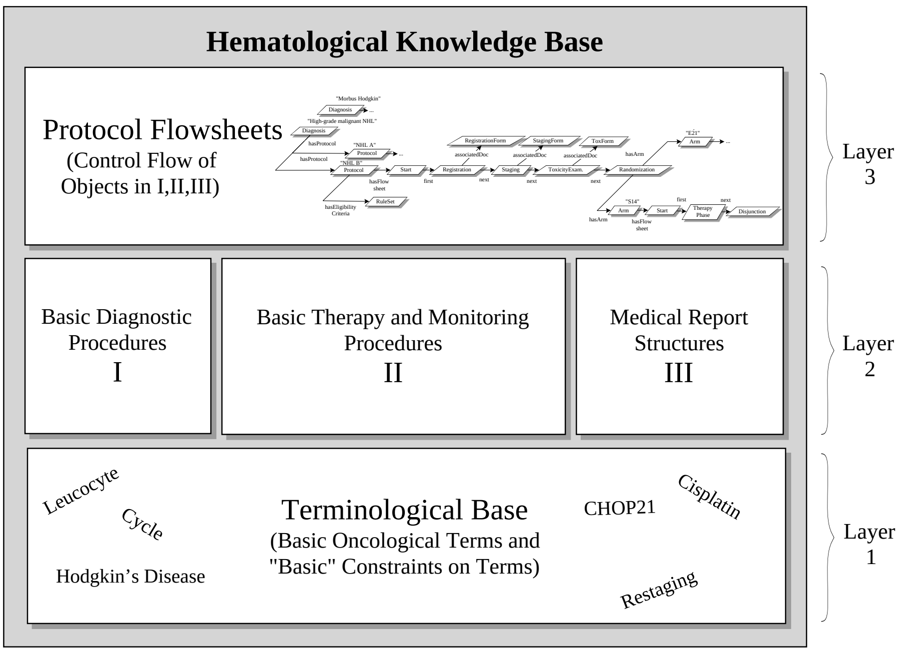


Figure 4: Structure of the oncological knowledge base.

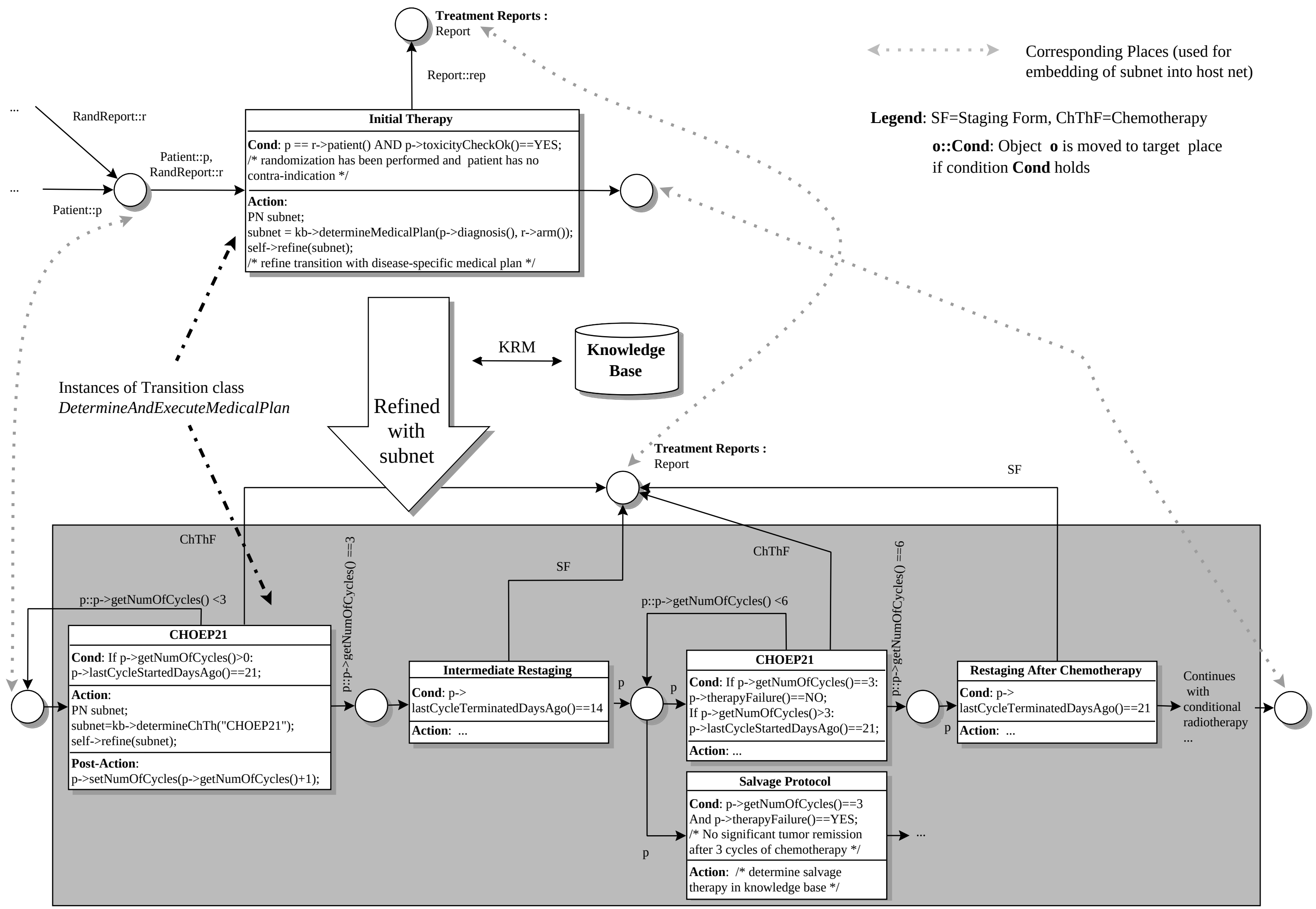
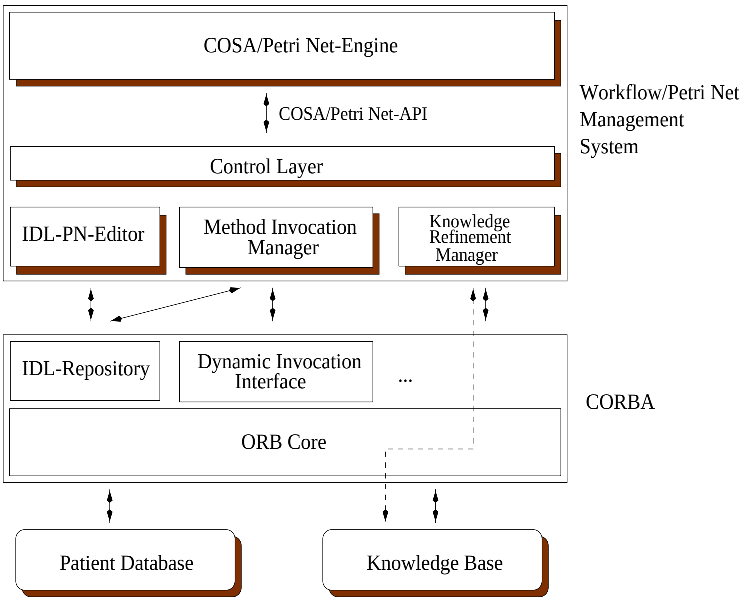


Figure 5: Refinement of the abstract transition named *Initial Therapy*, which is an instance of the transition class *DetermineAndExecuteMedicalPlan* and part of a subnet refining the abstract high-level transition *Treat Patient* of figure 3. In this figure, this transition is refined by a disease-specific net for the treatment of lymphoma tumor, which has been retrieved from the knowledge base and is dynamically build into the net by the *Workflow Refinement Manager* (see section 5). This initial treatment in the gray rectangle consists of 3 cycles of the chemotherapy of type CHOEP21 (an acronym from the first letters of the involved drugs), a diagnostic scanning (restaging) after 14 days, additional 3 cycles of chemotherapy, and a final diagnostic examination (restaging after the last chemotherapy). The transitions named "CHOEP21" are instances of the transition class *DetermineAndExecuteChemotherapy* (which is a subclass of *DetermineAndExecuteMedicalPlan*); "kb" in the transitions action part is a global CORBA reference to the knowledge base.

Figure 6: Basic components of the Workflow Management System, which is based on the PN-oriented COSA System.



communicating with the other workflow systems. Communication via the *Inter Hospital Communication*-places of figure 3 is realized with IONA ORBIXWEB. As the COSA WORKFLOW system does not directly support object-oriented PNs and is not CORBA-compliant, it is currently extended with the following modules (see figure 6):

- An PN-Editor communicating with the ORBIX IDL-Repository to define IDL-classes (of tokens, places and transitions) and to specify IDL-method calls in the PN-transitions. The IDL-compliant PN-definitions are then internally transformed to COSA nets (by adding ”technical” places and transitions for the IDL-method calls etc.). In a COSA net, object (string) references are represented by COSA string variables.
- As the object string references are meaningless to the COSA engine, an additional *Method Invocation Manager* has to be implemented. This manager, which communicates with the COSA kernel at run-time, handles the IDL-calls specified in the net definitions (by using the ORBIX DYNAMIC INVOCATION Interface) and sends the results back to COSA.
- Furthermore, the *Knowledge Refinement Manager*, which is mediating between the workflow engine and the domain knowledge base (see section 5) is implemented.
- For the communication between the COSA engine and the modules described above, a *Control Layer* completes the extensions.

Beside that, the following databases and applications, which are controlled by the workflow systems, have been implemented or are currently under development (every database listed is wrapped into IDL-CORBA objects using the Orbix Database Adapter Framework):

- An Oracle 7-based database application for oncological patient data [14]. Every access to a *Patient* token object during PN-execution establishes a communication with this database.
- A report editor and generator to define the structure of medical documents (of the class hierarchy of figure 1 for example) and to generate patient-specific reports sent to external departments. The report tools have been implemented in Personal Oracle 7 and Delphi 2.0 [15].
- An O_2 knowledge base for the disease-specific knowledge described in section 5. To implement the semantic networks mentioned in section 5, the abstract O2/C++ classes *Node*, *Edg*e and *Network* have been realized, which manage all network-related data accesses (such as adding and deleting nodes and edges to the network). From this classes, all disease-specific objects are derived.
- A tracking database at the central commission containing all report data received from the treating departments, and used by the tracking transition in figure 3 ([16]).

A critical point concerning implementation is that the approach described uses CORBA objects of a relatively fine granularity (e.g., *Report* and *Request* tokens), although CORBA principally has been designed for objects (services) of a larger granularity (e.g. [17][18]). This problem is partially resolved, as

- the number of patients treated simultaneously is relatively small (about 10 because of the restricted number of beds available at a hemato-

oncological site),
- queries over object collections are primarily processed within the databases themselves, and
- transaction management of the persistent token objects processed by the PN is realized by the ORBIX Database adapter used for every database and knowledge base within the ORBIX net.

However, this problem of using CORBA for "data-shipping" in the particular context of hematological data must be addressed further, including an application-specific realization of controlled object migration to reduce communication load.

7 Conclusion

This article described HEMATOWORK, a knowledge-based distributed workflow system for cancer therapy. The medical domain of hemato-oncology is characterized by geographically distributed sites involved in treatment (oncological ward, external expert panels and external central commissions), a rather complex treatment and complicated communication paths for the distribution of medical material and reports between the different sites. The modeling approach is based on CORBA-compliant, object-oriented, hierarchical PNs, using knowledge-based run-time refinement to refine treatment nets with disease-specific subnets. The implementation uses IONA ORBIX and the PN-oriented workflow management system Cosa, which is extended to manage the CORBA-related aspects of the used PN-language. Central problems, that have to be addressed in the future, include the efficient realization of "data-shipping"-aspects within the ORBIX net and the integration of temporal aspects.

References

[1] Taboada, M., Marín, R., Mira. J., Otero, R.P.: Integrating Medical Expert Systems, Patient Data-Bases and User Interfaces. *Journal of Intelligent Information Systems* 7 (3), 1996: 261-285.

[2] Ramirez, C. G. R., Smith, L.A, Patterson, L.L.: Medical Information Systems: Characterization and Challenges. Sigmod Record, 23 (3), 1994.

[3] Müller, R., Sergl, M., Nauerth, U. et al.: TheMPO: A Knowledge-Based System for Therapy Planning in Pediatric Oncology. *Computers in Biology and Medicine* vol. 27(3), 1997:177-200.

[4] Georgakopoulos, D., Hornick, M., Sheth, A.: An Overview of Workflow Management: From Process Modeling to Infrastructure for Automation. *Journal on Distributed and Parallel Database Systems* 3 (2), 1995: 119-153.

[5] Cichocki, A., Helal, A.S., Rusinkiewicz, M., et al..: Workflow and Process Automation: Concepts and Technology. *Kluwer Academic Publishers*, 1997.

[6] Sheth, A., Kochut, K., Miller, J. et al.: Supporting State-wide Immunization Tracking using Multi-Paradigm Workflow Technology. *Proceedings VLDB96*: 263-273.

[7] Lee, W., Kaiser, G.E.,Clayton, P.D., Sherman, E.H.: *OzCare: A Workflow Automation System for Care Plans. Proceedings of the 1996 American Medical Informatics Association Fall Symposium, Washington DC (Oct. 1996)*: 577-581.

[8] Dazzi, L., Fassino, C., Saracco, R. et al.: *A Patient Workflow Management System Built on Guidelines. Proceedings of the 1997 American Medical Informatics Association Fall Symposium, Nashville, TN (Oct. 1997)*: 146-150.

[9] Aalst, W.M.P. van der: Three Good reasons for Using a Petri-net-based Workflow Management System. In: Navathe, S. and Wakayama, T. (eds.): Information and Process Integration in Enterprises, *Kluwer Academic Publishers*, Norwell, 1997.

[10] Cosa Solutions: http://www.cosa.de/.

[11] Oberweis A., Scherrer, G., Stucky, W.: INCOME/STAR: Methodology and Tools for the Development of Distributed Information Systems. *Information Systems* 19(8), 1994: 643-660.

[12] Lakos, C.A.: From Coloured Petri Nets to Object Petri Nets. In: *Proceeding of the 16th International Conference on Application and Theory of Petri Nets, Turin, June 1995 (LNCS 935)*: 278-297.

[13] Schöf, S., Sonnenschein, M., Wieting, R.: Efficient Simulation of THOR Nets. *In: Proceeding of the 16th International Conference on Application and Theory of Petri Nets, Turin, June 1995 (LNCS 935)*: 412-431.

[14] Brümmer, F.: A Patient Database to Represent Oncological Treatments (in German). Master Thesis, University of Leipzig, 1997.

[15] Jödecke, E.: Concept und Implementation of an Editor for Document Structures. (in German). Master Thesis, University of Leipzig, 1997.

[16] Speer, R.: Concept and Realization of a Client/Server-Database for Clinical Trials (in German). Master Thesis, University of Leipzig, 1997.

[17] Dogac, A., Dengi, C., Kilic, E. et al.: A Multidatabase System Implementation on CORBA. *Proceedings of the Sixth International Workshop on Research Issues in Data Engineering - New Orleans, Louisiana, February 26-27, 1996*: 2-11

18. Sellentin, J., Mitschang, B.: Data-Intensive

Intra- & Internet Applications - Experiences Using Java and CORBA in the World Wide Web. *Proceedings of the 14th IEEE International Conference on Data Engineering, Orlando, Florida, February 1998*.