Towards Bounded-Rationality in Multi-Agent Systems: A Reinforcement-Learning Based Approach

Anita Raja and Victor Lesser

November 13, 2001

Abstract

Sophisticated agents operating in open environments must make complex real-time control decisions on scheduling and coordination of domain activities. These decisions are made in the context of limited resources and uncertainty about outcomes of activities. The question of how to sequence domain and control activities without consuming too many resources in the process, is the meta-level control problem for a resource-bounded rational agent. Our approach is to design and build a meta-level control framework with bounded computational overhead. This framework will support decisions on when to accept, delay or reject a new task, when it is appropriate to negotiate with another agent, whether to renegotiate when a negotiation task fails and how much effort to put into scheduling when reasoning about a new task.

1 Introduction

Sophisticated agents operating in complex environments must reason about their local problem solving activities, interact with other agents, plan a course of action and carry it out. All these have to be done in the face of limited resources and uncertainty about action outcomes in real-time. Furthermore, new tasks can be generated by existing or new agents at any time, thus an agent's deliberation must be interleaved with execution. The planning, scheduling and coordination of tasks are non-trivial, requiring either exponential work, or in practice, a sophisticated scheme that controls the complexity. In this paper, we describe a framework which will provide effective allocation of computation resulting in improved performance of individual agents in a cooperative multi-agent system.

In this framework, agent activities are broadly classified into three categories - **domain**, **control**, and **meta-level control** activities. Domain activities are executable primitive actions which achieve the various high-level tasks. Control activities are of two types, scheduling activities which choose the high level tasks, set constraints on how to achieve them and sequence the detailed domain level activities which achieve the selected tasks; and coordination activities which facilitate cooperation with other agents in order to achieve the high-level tasks. Meta-level control activities optimize the agent's performance by choosing and sequencing domain and control activities.

Agents perform control activities to improve their performance. Many efficient architectures and algorithms that support these activities have been developed and studied[1, 9, 11]. Agents receive sensations from the environment and respond by performing actions that affect the environment using the effectors. The agent chooses its domain level activities and this might involve invoking the scheduling and coordination modules. Classic agent architectures either overlook the cost of control activities or they assume a fixed and negligible cost and do not explicitly reason about the time and other resources consumed by control activities, which may in fact degrade an agent's performance. An agent is not performing rationally if it fails to account for the overhead of computing a solution. This leads to actions that are without operational significance [12].

Consider an administrative agent which is capable of multiple tasks such as answering the telephone, paying bills and writing reports. It usually takes the agent a significant amount of time to sort out the bills. Suppose the agent does not perform any meta-level reasoning about the importance or urgency of the tasks. It will then spend the same amount of time deciding whether to pick up a ringing phone as it does on deciding which bills to pay. If the agent is equipped with meta-level reasoning capabilities, it will recognize the need to make quicker decisions on whether to answer the phone than on sorting bills since there is external constraint on the ringing phone, namely that the caller could hang up. The agent will make better decisions on answering calls as well as completing its other tasks by dynamically adjusting its decision based on its current state and the incoming task.

Figure 1 describes our architecture which will support this dynamic adjustment process by introducing resource-bounded meta-level reasoning in agent control. The classic architecture is augmented with a meta-level control component and there are various options for invoking the scheduling and coordination components. These options differ in their resource usage and performance. The meta-level control component will decide if, when and how much control activity is necessary for each event sensed by the agent.

Meta-level control activities include allocating appropriate amount of processor and other resources at appropriate times. To do this an agent would have to know the effect of all combinations of actions ahead of time, which is intractable for any reasonably sized problem. The question of how to approximate this ideal of sequencing domain and control activities without consuming too many resources in the process, is the **meta-level control problem** for a resource bounded rational agent. In this paper, the approximation is done using a case-base of hand-generated heuristics which are described in detail in Section 3. The assumptions made in our solution approach are enumerated in Section 2. In Section 5, we provide a review of meta-level control research. Experimental results illustrating the strength of meta-level control in agent reasoning are discussed in Section 4.

2 Assumptions

The following assumptions are made in the framework described in this paper: The agents are cooperative and will prefer alternatives which increase social utility/quality even if it is at the cost of decreasing local utility. An agent may concurrently pursue multiple high-level goals and the completion of a goal derives quality for the system or agent. The

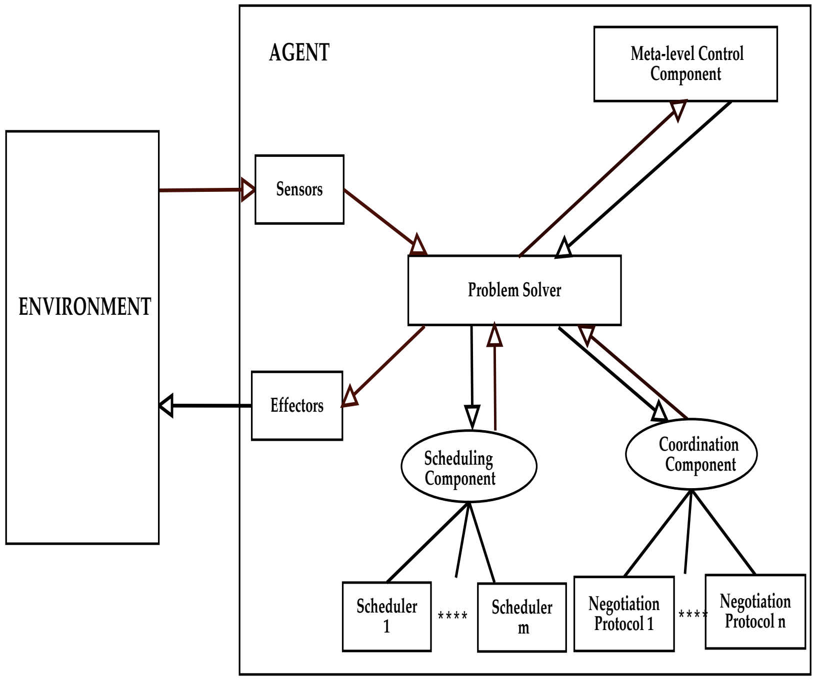

Figure 1: New architecture of a bounded rational agent

overall goal of the system or agent is to maximize the quality generated over some finite time horizon. The high-level goals are generated by either internal or external events being sensed and/or requests by other agents for assistance. These goals must often be completed by a certain time in order to achieve any quality. It is not necessary for all high-level goals to be completed in order for an agent to derive quality from its activities. The partial satisfaction of a high-level goal is sometimes permissible while trading-off the amount of quality derived for decrease in resource usage. The agent's scheduling decisions involve choosing which of these high-level goals to pursue and how to go about achieving them. There can be non-local and local dependencies between tasks and methods. Local dependencies are inter-agent while non-local dependencies are intra-agent. These dependencies can be hard or soft precedence relationships. Coordination decisions involve choosing the tasks which require coordination and also which agent to coordinate with and how much effort much be spent on coordination. Scheduling and coordination activities do not have to be done immediately after there are requests for them and in some cases may not be done at all. There are alternative ways of completing scheduling and coordination activities which trade-off the likelihood of these activities resulting in optimal decisions versus the amount of resources used. We also make the simplifying assumption that negotiation results are binding and we assume that the agents will not decommit from their contract at later stages.

3 Agent Architecture

In this section, we provide an overview of our architecture which provides effective meta-level control for bounded rational agents. Figure 2 describes the control flow within this proposed architecture. The number sequences describe the steps in a single flow of control. At the heart of the system is the **Domain Problem Solver**(DPS). It receives tasks and other external requests from the environment(Step **1**). When an exogenous event such as *arrival of a new task* occurs, the DPS sends the corresponding task set, resource constraints as well constraints of other tasks which are being executed, and performance criteria to the meta-level controller(Step **2**). The controller computes the corresponding state and determines the best action prescribed by the hand-generated heuristic policy for that particular task

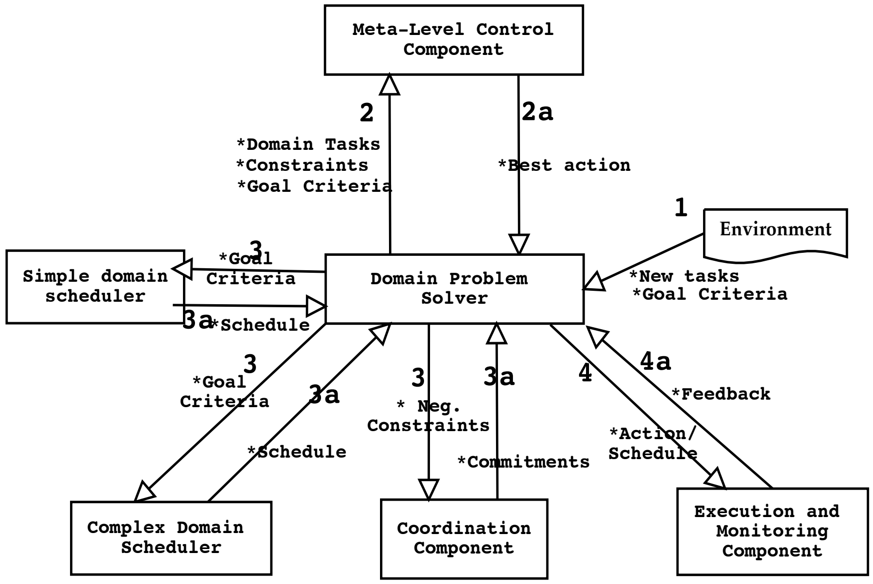

Figure 2: Control-flow in a bounded rational agent

environment. The best action can be one of the following: to call one of the two domain schedulers on a subset of tasks; to gather more information to support the decision process; to drop the new task or to do nothing. The meta-level controller then sends the prescribed best action back to the DPS(Step **2a**).

The DPS, based on the exact nature of the prescribed action, can invoke the **complex scheduler**, **simple scheduler** or **coordination component**(Step **3**) and receives the appropriate output(Step **3a**). If the action is to invoke the complex scheduler, the scheduler component receives the task structure and objective criteria as input and outputs the best satisficing schedule as a sequence of primitive actions. The complex scheduler can also be called to determine the constraints on which a coordination commitment is established. If the meta-level or the domain scheduler prescribe an action that requires establishing a commitment with a non-local agent, then the coordination component is invoked. The coordination component receives a vector of commitments that have to be established and outputs the status of the commitments after coordination completes. The simple scheduler is invoked by the DPS and receives the task structure and goal criteria. It uses pre-computed abstract information of the task to select the appropriate schedule which fits the criteria.

The DPS can invoke the execution component either to execute a single action prescribed by the meta-level controller or a schedule prescribed by the domain-level scheduler(Step **4**). The execution results are sent back to the DPS(Step **4a**) where they are evaluated and if the execution performance deviates from expected performance, the necessary measures are taken by the DPS.

This work accounts for the cost at all three levels of the decision hierarchy - domain, control and meta-level control activities. The cost of domain activities is modeled directly in the task structures which describe the tasks. The cost of domain activities are reasoned about by control activities like negotiation and scheduling.

The cost of control activities are reasoned about by the meta-level control activities. Negotiation costs are reasoned about explicitly in this framework since they can be modeled as part of the domain activities needed to complete a high-level goal. The negotiation tasks are split into an information gathering phase and a negotiating phase, with the outcome of the former enabling the latter. The negotiation phase can be achieved by choosing a member from a family of negotiation protocols[15]. The information gathering phase is modeled as a **MetaNeg** method in the task structure and the negotiation methods are modeled as individual primitive actions. Thus, reasoning about the costs of negotiation is done explicitly, just as it is done for regular domain-level activities. The **MetaNeg** method belongs

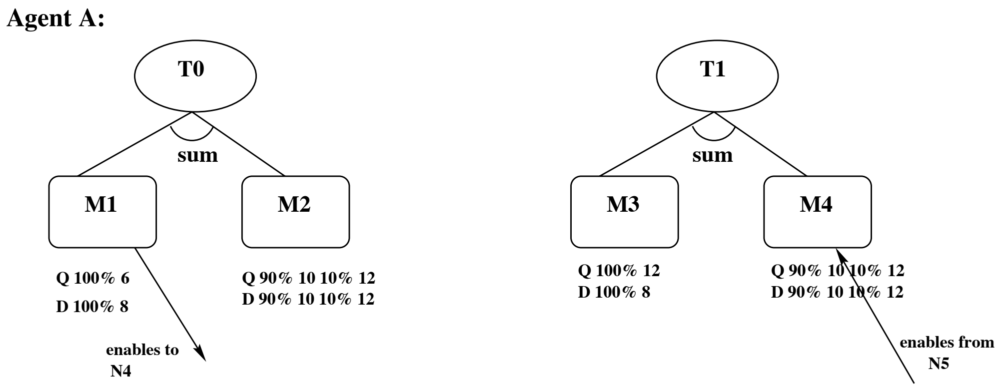

Figure 3: Tasks that can be performed by agent A

to a special class of domain actions which request an external agent for a certain set of information and it does not use local processor time. It queries the other agent and returns information on the agent's expected quality from its tasks, expected completion time of its tasks and flexibility of its schedule. This information is used by the meta-level controller to determine the relevant control actions.

However, reasoning about the cost associated with scheduling activities is implicit. A fixed cost is associated with each of the two schedulers and these costs affect the subsequent choice of domain activities made by the control activities. The earliest start time of domain activities are determined by the latest finish times of their corresponding control activities.

Meta-level control activities in this framework are modeled as inexpensive activities. The cost for meta-level control in this framework are incurred by the computation of state features which facilitate the heuristic decision-making process. The state features and their functionality are described in greater detail later on in this section.

The domain level scheduler depicted in the architecture will be an extended version of the Design-to-Criteria(DTC) scheduler[14]. Design-to-Criteria (DTC) scheduling is the soft real-time process of finding an execution path through a hierarchical task network such that the resultant schedule meets certain design criteria, such as real-time deadlines, cost limits, and quality preferences. It is the heart of agent control in agent-based systems such as the resource-Bounded Information Gathering agent BIG [8]. Casting the language into an action-selecting-sequencing problem, the process is to select a subset of primitive actions from a set of candidate actions, and sequence them, so that the end result is an end-to-end schedule of an agent's activities that meets situation specific design criteria.

We also introduce a simple scheduler based on the use of abstractions of agent task structures. This will support reactive control for highly constrained situations. Abstraction is an offline process where potential schedules and their associated performance characteristics for achieving the high level tasks are discovered for varying objective criteria. This is achieved by systematically searching over the space of objective criteria. Also multiple schedules could potentially be represented by the same abstraction. The abstraction hides the details of these potential schedules and provides only the high level information necessary to make meta-level choices. When an agent has to schedule a task but doesn't have the resources or time to call the complex domain-level scheduler, the generic abstraction information of the task structure can be used to provide the approximate schedule.

Taxonomy of meta-level control decisions

We now describe a taxonomy of the meta-level decisions in a multi-agent system using a simple example scenario. Consider a multi-agent system consisting of 2 agents *A* and *B*. The discussion will focus only on the various meta-level questions that will have to be addressed by agent *A*. Figure 3 describes *T0* and *T1*, which are the tasks performed by agent *A*. They are described using TÆMS, a domain independent framework for describing task structures.

In this example, each top-level task is decomposed into two executable primitive actions. In order to achieve the task, agent *A* can execute one or both of its primitive actions within the task deadline and the quality accrued for the task will be cumulative (denoted by the *sum* function). Methods are primitive actions which can be scheduled and executed and are characterized by their expected quality, cost and duration distributions. For instance, the quality distribution of method *M2* indicates that it achieves quality value of 10 with probability 0.9 and quality of 12 with probability 0.1. Quality is a deliberately abstract domain-dependent concept that describes the contribution of a particular action to overall problem solving. The *enables* relationship from method *M1* of task *T0* to a non-local method *N4* of task *S1* belonging to agent *B* (agent *B's* task structure is not shown) implies that successful execution of *M1* is a precondition for executing *N4*.

In the remainder of this section, we enumerate the features computed when the meta-level control component is invoked. The cost of computing and reasoning about these state features reflect the cost of meta-level control reasoning. We then enumerate the various meta-level control decisions and the case-base of heuristics used to make the decisions.

The following are some simple state features which are used in the heuristic decision making process of the meta-level controller.

F0: Current status of system This feature is represented as a 3-tuple $< NewItemsStack, Agenda, ScheduleStack >$ where each entry in the tuple contains the number of items on the corresponding stack. The new items are the tasks which have just arrived at the agent from the environment. The agenda stack is the set of tasks which have arrived at the agent but whose reasoning has been delayed and they have not been scheduled yet. The schedule stack is the set of tasks currently being scheduled. Eg. $< 2, 0, 1 >$ means there are two new items which have arrived from the environment and there is one task being scheduled.

F1: Relation of quality gain per unit time of a particular task to that of currently scheduled task set: The $utility\ gain\ per\ unit\ time$ of a task is the ratio of $total\ expected\ utility$ to $total\ expected\ duration$ of that task. This feature compares the utility of a particular task to that of the existing task set and helps determine whether the new task is very valuable, moderately valuable or not valuable in terms of utility to the local agent.

F2: Relation of deadline of a particular task to that of currently scheduled task set: This feature compares the deadline of a particular task to that of the existing task set and helps determine whether the new task's deadline is very close, moderately close or far in the future.

F3: Relation of priority of items on agenda to that of currently scheduled task set: This feature compares the average priority of the existing task set to the priority of the new task and helps determine whether the new task is very valuable, moderately valuable or not valuable in terms of utility to the local agent. Priority is a function of the utility and deadlines of the tasks. Computing the average priority of a task set is a more complicated function than computing the priority of a single tasks since it involves recognizing dominance of individual tasks.

The following are some of the specific meta-level issues that will be addressed by any individual agent.

1. Arrival of a new task from the environment: When a new task arrives at the agent, the meta-level control component has to decide whether to reason about it later; drop the task completely; or to do scheduling-related reasoning about an incoming task at arrival time and if so, what type of scheduling - complex or simple.
   Heuristic Rule: If the new task has very low or negligible priority and high opportunity cost with respect to taking resources away from future higher priority tasks, then it should be discarded. If the incoming task has very high priority, in other words, the expected utility is very high and it has a relatively close deadline, then the agent should override its current schedule and schedule the new task immediately. If the deadline is very tight the agent will uses the abstraction-based simple scheduler; else, it will use the more complex scheduler. If the current schedule has average utility that is significantly higher than the new task and the average deadline of the current schedule is significantly closer than that of the new task, then reasoning about the new task should be postponed till later. If the new task is scheduled immediately, the scheduling action costs time, and there are associated costs of dropping established commitments if the previous schedule is significantly revised or completely dropped. These costs are diminished or avoided completely if the task reasoning is postponed to later or completely avoided if the task is dropped.

2. Decision on whether to negotiate: The meta level controller will decide to negotiate based on the information returned by the **MetaNeg** action. It queries the other agent and returns information on the agent's expected quality from its tasks, expected completion time of its tasks and flexibility of its schedule. In Figure 3, method *M4* in agent *A* is enabled by method *N5* belonging to agent *B*. The benefit from including method *M4* in agent

A's schedule is that it increases its total utility. However, it also requires agent *A* and *B* to negotiate over the completion time of method *N5* by agent *B* and this negotiation has an associated cost as well as there is a resource cost to the agent which agrees to the contract
Heuristic Rule: If the other agent's current expected utility is much lower than the results of the negotiation, then the local agent will initiate negotiation. Negotiation is also initiated if the other agent's tasks have high utility but the deadlines are far enough in the future to permit the other agent to execute the enabling task. If the other agent's tasks have higher priority than the local task, then the negotiation option is dropped.

3. Choice of negotiation protocol: When an agent decides to negotiate, it should also decide whether to negotiate by means of a single step or a multi-step protocol that may require a number of negotiation cycles to find an acceptable solution or even a more expensive search for a near-optimal solution. The single shot protocol is quick but has a higher chance of failure where as a more complex protocol takes more time and has a higher chance of success.
Heuristic Rule: If the agent receives high utility from the results of the negotiation, then the agent should choose the more effective albeit more expensive protocol. The protocol which has a higher guarantee of success require more resources, more cycles and more end-to-end time in case of multi-step negotiation and higher computation power and time in case of near-optimal solutions. (The end-to-end time is proportional to the delay in being able to start task executions). If the agent does not have too much resources to expend on the negotiation or if there is a very slight probability that the other agent will accept the contract, then the local agent should choose the single shot protocol.

4. Failure of a negotiation to reach a commitment If the negotiation between two agents using a particular negotiation protocol fails, the initiating agent should decide whether to retry the negotiation; whether to use the same protocol or an alternate protocol with the same agent or alternate agents and how many such retries should take place?
Heuristic Rule: If negotiation is preferred (the agent will receive high utility as result of the negotiation), then a more complex negotiation protocol is chosen since it has a higher probability of succeeding. Since resources have already been spent on figuring out a solution to the negotiation, it may be profitable to put in a little more effort and achieve a solution. If there is a very slight or no probability of finding an acceptable commitment, then resources which can be profitably spent on other solution paths are being wasted and the agent might find itself in a dead-end situation with no resources left for an alternate solution. So the negotiation option should be dropped.

4 Experimental Results

Effective meta-level control is a complex process. It involves taking into account a number of factors, including task relationships, deadlines, the availability of alternatives, client design criteria (i.e., quality, cost, duration, and certainty trade-offs) and the execution state of other agents in the multi-agent system. In this section, we evaluate the performance of the meta-level control enhanced agent architecture by comparing it to the standard agent architecture which does not do explicit meta-level reasoning. It is possible to characterize the types of task environments that are amenable to meta-level reasoning, i.e., those for which dynamic allocation of appropriate amount of resources to domain and control activities at appropriate times is beneficial from a cost/benefit perspective. As part of the evaluation process, we have partially determined the characteristics of tasks, arrival models and design criteria that indicate a problem instance for which explicit meta-level control is advantageous. The general characteristics include:

1. Tasks arrive dynamically at the agent and have non-deterministic execution characteristics. If the agent knows the exact task arrival model as well as execution characteristics ahead of time, then real-time meta-level control is dispensable since a near-optimal policy can be built offline.
2. There are alternative ways of completing scheduling and coordination activities which trade-off the likelihood of these activities resulting in optimal decisions versus the amount of resources used. The meta-level controller will be able to dynamically adjust the computational resources only if such alternatives are available.

Row #	Agent Name	TS ID	Arrival Time	Deadline	Control Activity	Quality						
						Run1	Run2	Run3	Run4	Run5	Run6	Run7
1	A	T0_1	1	40	NTCS	17.50	19.50	17.50	17.50	17.50	17.50	16.30
2	A	T0_2	10	28	Drop	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	A	T1_3	21	75	ATCS, NM1	45.60	51.19	55.20	45.60	45.60	12.00	43.41
4	A	T0_4	55	80	ATCS	16.00	16.00	16.00	16.00	16.00	16.00	18.00
5	A	T0_5	61	100	ATCS	6.00	6.00	6.00	6.00	6.00	6.00	6.00
6	B	NX_3	37	47	SS	10.00	10.00	10.00	10.00	10.00	-	10.00
Total Quality						96.10	102.69	104.70	95.10	95.10	51.50	93.77
Col. #	1	2	3	4	5	6	7	8	9	10	11	12

Table 1: Experimental Results for agents capable of meta-level reasoning: *A*gent Name is name of the agent being considered; *T*S ID is the name of the task being considered; *A*rrival Time and *D*eadline are the arrival times and deadlines for that particular task; *C*ontrol Activity is the control action chosen by the meta-level controller; Columns 5-11 describe the quality accrued for each of the individual tasks in seven different runs; Row 6 describes the total quality of all tasks completed by both agents for each run

3. Methods should be interruptible. When one of the five exogenous events described in Section 3 occurs, the system state is immediately saved, execution of primitive actions is halted and control is shifted to the meta-level controller.
4. The system state as defined by feature **F0** should be accessible to the local agent. When cooperation with other agents is necessary, the local agent should have access to high level information on the status of other agents via the **MetaNeg** method.

The following are the design criteria characteristics which augment meta-level control.

1. The objective function should be to maximize overall quality over a given finite horizon. The hard deadline and other such hard resource constraints voids the possibility of simply rescheduling at failure points or initiating negotiation exchanges which are bound to fail. They instead require the high-level analysis of the meta-level control component which will decide whether it is worth while to invest the resources into such control actions.
2. The deadline should also provide enough time for control actions, since the scheduling and negotiation costs are factored into the equation. If the deadlines are too tight, then there can be no reasonable way of performing the relevant control actions and achieving quality for the tasks.

The experiments in this section are preliminary results to establish the plausibility of the theory described in this paper. Further experiments are in progress. For the purposes of this paper, we selected a randomly generated environment which adheres to the above mentioned characteristics. This produced tasks and arrival models amenable to meta-level control. The particular environment described here is simple and consists of two agents Agent *A* and Agent *B*. There are three possible tasks in the environment - task **T0**, task **T1** and task **NX**. The former two tasks can only be performed by agent *A* and task **NX** can only be performed by agent *B*. Task **T1** requires a non-local enablement from task **NX**. The maximum possible quality from task **T0** is 23.00 and minimum is 17.00; the maximum from task **T1** is 56.00 and minimum quality is 12.00, **NX** has a deterministic quality of 10.00. We make the simplifying assumption that task **NX** arrives at agent *B* only as a result of a successful negotiation with agent *A*. There are four possible meta-decisions upon arrival of a new task: **NTCS**, New Task Complex Scheduling invokes the complex DTC scheduler on the new task only and has a time cost of 2; **Drop**, this causes the agent to drop the new task and not reason about it ever again has a time cost of 0; **ATCS**, All Task Complex Scheduling invokes the complex DTC scheduler on the new task as well as all other tasks which are on the agenda or in partial execution and has a time cost of 3; and **SS**, Simple Scheduling invokes the simple abstraction based analysis on the new task only and has a time cost of 1. There are two possible options for Negotiation: **NM1**, Negotiation Mechanism 1 which is the simple single-shot protocol and **NM2**, Negotiation Mechanism 2 which is the more complex multi-shot protocol.

Row #	Agent Name	TS ID	Arrival Time	Deadline	Control Activity	Quality						
						Run1	Run2	Run3	Run4	Run5	Run6	Run7
1	A	T0_1	1	40	ATCS	22.80	24.00	23.00	22.00	22.00	22.00	18.26
2	A	T0_2	10	28	ATCS	10.00	12.00	10.00	10.00	10.00	10.00	10.00
3	A	T1_3	21	75	ATCS, NM1	12.00	12.00	12.00	12.00	12.00	12.00	12.00
4	A	T0_4	55	80	ATCS	10.00	10.00	10.00	10.00	10.00	10.00	12.00
5	A	T0_5	61	100	ATCS	10.00	10.00	12.00	12.00	10.00	10.00	10.00
6	B	NX_3	39	53	ATCS	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Total Quality						74.80	78.00	77.00	76.00	74.00	74.00	72.26
Col. #	1	2	3	4	5	6	7	8	9	10	11	12

Table 2: Experimental Results for agents with no meta-level reasoning: *C*ontrol Activity is the fixed control action used by the agent

The design criteria in these experiments is to maximize overall quality over a finite horizon. Individual tasks have hard deadlines associated with them. It is assumed that if a task has not accrued quality by its deadline, it receives a quality of zero. This simple design criteria setting is one that lends itself to meta-level control as the existence of a hard deadlines (in contrast to a soft preference, e.g., soft deadline or no deadlines) make processor and other resources valuable commodities requiring a the non-myopic reasoning provided by the meta-level control component.

The results for the experiments on agents which have meta-reasoning capabilities are shown in Table 1 and the results on agents which have no meta-level reasoning capabilities are shown in Table 2. The above described scenario is used in both cases. All domain, control and meta-level actions have a time cost associated with them which are reflected in the results.

Consider Table 1. Each row in the table represents a specific task arriving at the specified agent at the associated arrival time with a deadline. The task names are augmented with the arrival count to differentiate between various instances of the same task. For eg. Row 4 describes task **TO** arriving at agent *A* as its fourth task at time 55 with a deadline of 80. Column 5, titled *Control Action* describes the various decisions made by the meta-level controller upon arrival of the new task. Columns 6-12 describe the quality accumulated by each of the tasks for seven different runs.

In Row 1,task **T0_1** arrives at time 1. Since the meta-level controller is aware that no other tasks are in execution, it invokes **NTCS** on the task which is a cheaper option than **ATCS** which would be the choice of an agent with no meta-reasoning capabilities.

In Row 2, task **T0_2** arrives at time 11 while the previous task is still in execution and a meta-level decision to drop task **T0_2** is made. This is because the previous task **T0_1** has the exact same characteristics as the current task and has a tight deadline. The task also has a tight deadline and interrupting the already executing tasks might result in missing the deadlines on both Task **T0_1** and task **T0_2**.

In Row 3, Agent *A* decides to do a complete reschedule of all tasks and chooses to negotiate with agent *B* over task **NX** using negotiation mechanism **NM1**. In this case, it is willing to reschedule task **T0_1** since the expected quality from the newly arrived task is much higher than that of the current task. Also, the fact that the agent dropped task **T0 2** although it was unaware of the arrival of a highly preferred task in the near-future works to agent *A's* advantage since it has more time to perform the higher valued task. In five out of six runs, the agent's decision to drop the previous task **T0_2** and perform task **T1_3** with the negotiation option results in very high quality values. In Run 6, task **T1 3** receives a very low quality because negotiation fails with agent *B* and the task receives the minimum quality. However on average, agent *A*'s meta-level decision works to its benefit.

In Row 6, we see that agent B̂ chooses the simple scheduling option to execute task **NX_3** because of its tight deadline.

Consider Table 2. Here the agent does not reason about the characteristics of the tasks at the meta-level. This results in the agent choosing the same control action, namely **ATCS** for all tasks independent of the status of other tasks in execution. This results in the most expensive control action being invoked independent of the current state of the system. This results the choice of domain activities with shorter durations and lower qualities as reflected by the quality values in columns 6-12.

The total qualities accumulated by five of the six runs in Table 2 is significantly lower than the corresponding run in Table 1. This supports our hypothesis that meta-level control is generally advantageous.

5 Related Work

Meta-level control has also been called meta-level planning [13]. As this term implies, an agent can plan not only the physical actions that it will take but also the computational actions that it will take. The method for performing this planning can range from simple heuristics to recursive application of the full planner. Stefik's Molgen planner uses the base level planner to create meta-level plans. Molgen considers two levels of meta-level planning, in addition to base-level planning. The actions at each of these meta-levels create plans for the next lower level. In contrast, our approach uses only a single layer of meta-level control and uses algorithms and heuristics tailored to making particular meta-level control decisions. Additional layers of meta-level control have a diminishing rate of return since each layer adds additional overhead and there is a limit on how much meta-level control can improve performance.

In order to make the tradeoffs necessary for effective meta-level control, the meta-level controller needs some method for predicting the effect of more computation on the quality of a plan. One method for doing this is to use a performance profile. The idea comes from the study of anytime algorithms that can be interrupted at any point to return a plan that improves with more computation [2]. The performance curve gives the expected improvement in a plan as a function of computation time. Anytime algorithms can also be combined to solve complex problems. Zilberstein and Russell look at methods for combining anytime algorithms and performing meta-level control based on multiple performance curves [16]. Combining anytime algorithms produces new planning algorithms that are also characterized by a performance curve.

[4] extends previous work on meta-level control of anytime algorithms by using a non-myopic stopping rule. It finds an intermediate strategy between continuous monitoring and not monitoring at all. It can recognize whether or not monitoring is cost-effective, and when it is, it can adjust the frequency of monitoring to optimize utility. This work has significant overlap with the foundations of the approach described in this paper. However, it is in a multi-agent non-anytime setup with interacting agents which makes the decision-making process more complex.

The partial global planning [3] approach is a flexible framework for coordination where nodes can balance their needs for predictability and responsiveness differently for different situations. In this framework, nodes exchange information about their tentative local plans and develop partial global plans(PGPs) to represent the combined activities of some part of the network that is developing a more global solution. To dampen their reactions to deviations, nodes need to know when deviations are negligible and should be ignored. The PGPlanner considers a deviation between actual and predicted times to be negligible if that difference is no larger than the time-cushion. The time-cushion is a user-specified parameter that represents negligible time and balances predictability and responsiveness. In the work presented here, we make similar decisions on predictability and responsiveness. Our approach is more general since we do not have preset parameters to handle each of the decisions. The agent can dynamically adjust its response based on its current state.

[10] presents a learning system called COLLAGE that uses meta-level information in the form of abstract characterization of the coordination problem instance to learn to choose the appropriate coordination strategy from among a class of strategies. They provide empirical evidence for the benefits of learning situation-specific coordination. [7] proposes a meta-level control mechanism for coordination protocols in a multi-agent system. AgenTalk, a coordination protocol description language, is extended to include primitives for the meta-level control. The meta-level control mechanism allows agents to detect and handle unexpected situations by switching between coordination protocols. These two systems deal with similar issues as this work. They choose a situation-specific strategy from a number of options. However they do not account for the cost of meta-level control. They also limit their work to coordination protocols and don't consider control activities.

An opportunistic control model that can support different control modes expected of an intelligent agent with multiple goals, limited resources, and dynamic environments is described in [5]. Their goal is similar to ours in that they are concerned with the fact that in dynamic environments, it is often necessary to make decisions that may not be optimal, but satisfactory under the current conditions. [6] discuss an experimental intelligent agent called Guardian for monitoring patients in a surgical ICU. It is based on the BB1 blackboard architecture developed by Hayes-Roth. Advantages claimed over traditional patient monitoring systems include multiple reasoning skills and the ability to operate under time pressure. The meta-level controller in GUARDIAN controls the amount of information fed to the input buffer at varying rates depending on the current situation.

6 Conclusions and Future Work

In this paper we present a novel meta-level control agent framework for sophisticated multi-agent environments. The meta-level control has limited and bounded computational overhead and will support reasoning about scheduling and coordination costs as first-class objects.

We have shown in Section 4, using a simple example, that meta-level control is beneficial. The heuristics described in this paper, although very simple, enable the meta-level controller to make accurate decisions in simple scenarios. We plan to introduce more complex features which will make the reasoning process more robust. Some such features include relation of slack fragments in local schedule to new task. This would enable an agent to fit a new task in its current schedule if it is possible and avoid a reschedule. Another feature would be to estimate the decommitment cost for a particular task. This will enable us to consider environments in which agents can decommit from tasks which they have previously agreed to complete.

We will be extending the detailed domain level scheduler(DTC) to handle scheduling effort, slack and horizon as first-class objects. The extended DTC will accept parameters which constrain the effort spent on scheduling which in turn will affect the overhead of the scheduler. It will also be extended to deal with slack as a schedulable element which can be quantified and valued as any other primitive action. We hope that augmenting the domain level scheduler will provide the meta-level controller with more options, hence making it more versatile.

We also plan to compare our heuristic based meta-level controller to a quasi-optimal policy which is built assuming there is complete knowledge of performance characteristics of the actions and exogenous events ahead of time for a specific scenario. The quasi-optimal policy will provide a performance upper-bound to our system and will help us study the strengths and weaknesses of our approach..

7 Acknowledgments

We thank Bryan Horling for his assistance in extending the JAF framework and Mass simulator to satisfy the requirements of the meta-level control component.

References

[1] Craig Boutlier. Sequential Optimality and Coordination in Multiagent Systems. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, 1999.

[2] Thomas Dean and Mark Boddy. An analysis of time-dependent planning. In *Proceedings of the Seventh National Conference on Artificial Intelligence (AAAI-88)*, pages 49–54, Saint Paul, Minnesota, USA, 1988. AAAI Press/MIT Press.

[3] E. Durfee and V. Lesser. Predictability vs. responsiveness: Coordinating problem solvers in dynamic domains. In *Proceedings of the Fifth National Conference on Artificial Intelligence*, pages 66–71, St. Paul, Minnesota, August 1988.

[4] Eric A. Hansen and Shlomo Zilberstein. Monitoring anytime algorithms. *SIGART Bulletin*, 7(2):28–33, 1996.

[5] B. Hayes-Roth. Opportunistic control of action in intelligent agents. In *IEEE Transactions on Systems, Man and Cybernetics*, pages SMC–23(6):1575–1587, 1993.

[6] B. Hayes-Roth, S. Uckun, J.E. Larsson, D. Gaba, J. Barr, and J. Chien. Guardian: A prototype intelligent agent for intensive-care monitoring. In *In Proceedings of the National Conference on Artificial Intelligence*, pages 1503–1511, 1994.

[7] Kazuhiro Kuwabara. Meta-level control of coordination protocols. In *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS96)*, pages 104–111, 1996.

[8] Victor Lesser, Bryan Horling, Frank Klassner, Anita Raja, Thomas Wagner, and Shelley XQ. Zhang. BIG:an agent for resource-bounded information gathering and decision making. In *Artificial Intelligence Journal, Special Issue on INternet Applications*, 1999.

[9] David J. Musliner. Plan Execution in Mission-Critical Domains. In *Working Notes of the AAAI Fall Symposium on Plan Execution - Problems and Issues*, 1996.

[10] M V Nagendra Prasad and Victor Lesser. The use of meta-level information in learning situation specific coordination. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence*, 1997.

[11] Anita Raja, Victor Lesser, and Thomas Wagner. Toward Robust Agent Control in Open Environments. In *Proceedings of the Fourth International Conference on Autonomous Agents*, pages 84–91, Barcelona, Catalonia, Spain, July,. ACM Press.

[12] H. Simon. From substantive to procedural rationality, 1976.

[13] M. Stefik. Planning and meta-planning, 1981.

[14] Thomas Wagner, Alan Garvey, and Victor Lesser. Criteria-Directed Heuristic Task Scheduling. *International Journal of Approximate Reasoning, Special Issue on Scheduling*, 19(1-2):91–118, 1998. A version also available as UMASS CS TR-97-59.

[15] XiaoQin Zhang, Rodion Podorozhny, and Victor Lesser. Cooperative, multistep negotiation over a multi-dimensional utility function. In *IASTED International Conference, Artificial Intelligence and Soft Computing (ASC 2000), Banff,Canada*, pages 136–142. IASTED/ACTA Press, July 2000.

[16] Shlomo Zilberstein and Stuart J. Russell. Efficient resource-bounded reasoning in AT-RALPH. In James Hendler, editor, *Artificial Intelligence Planning Systems: Proceedings of the First International Conference (AIPS 92)*, pages 260–268, College Park, Maryland, USA, 1992. Morgan Kaufmann.