

Approximate and Compensate: A method for risk-sensitive meta-deliberation and continual computation

David C. Parkes

Division of Engineering and Applied Sciences
Harvard University
Cambridge, MA 02138

Lloyd G. Greenwald

Mathematics and Computer Science
Drexel University
Philadelphia, PA 19104
lgreenwald@mcs.drexel.edu

Abstract

We present a flexible procedure for a resource-bounded agent to allocate limited computational resources to on-line problem solving. Our APPROXIMATE AND COMPENSATE methodology extends a well-known greedy time-slicing approach to conditions in which performance profiles may be non-concave and there is uncertainty in the environment and/or problem-solving procedures of an agent. With this method, the agent first approximates problem-solving performance and problem parameters with standard parameterized models. Second, the agent computes a risk-management factor that compensates for the risk inherent in the approximation. The risk-management factor represents a mean-variance tradeoff that may be derived optimally off-line using any available information. Theoretical and experimental results demonstrate that APPROXIMATE AND COMPENSATE extends existing methods to new problems and expands the practical application of meta-deliberation.

INTRODUCTION

We are interested in agents that actively manage how their limited computational resources are allocated to problem solving in time-constrained environments. This *meta-deliberation* problem can take the form of the agent allocating resources among a competing collection of problem-solving procedures (Georgeff & Lansky 1987; Garvey & Lesser 1993; Gomes & Selman 2001), among a sequence of future computational challenges (Boddy & Dean 1994; Zilberstein 1993), or among a competing collection of possible future challenges (Horvitz 2001; Greenwald & Dean 1994). A sample of domains for which meta-deliberation solutions have been suggested include Bayesian reasoning (Horvitz, Suermondt, & Cooper 1989; Horvitz & Breese 1990), robotics (Boddy & Dean 1994; Zilberstein & Russell 1993), graphics rendering (Horvitz & Lengyel 1997), and real-time scheduling (Greenwald & Dean 1994).

In order to make meta-deliberation decisions the agent must have some predictive information about the challenges (problem instances) it will encounter and problem-solving procedures available to apply to each projected challenge. A common example is a profile representing the expected quality of a result as a function of the amount of time spent

problem solving (Boddy & Dean 1994; Zilberstein & Russell 1996; Horvitz 2001).

For example, a time-critical medical decision support system must trade off the benefits of problem solving during decision support against the cost of delayed treatment. Every minute delayed in treating serious injury in an emergency center can reduce the probability of patient survival by approximately 1% (Clarke *et al.* 2001). While a surgeon is operating, the system might reason about possible future computational challenges, making use of probabilistic models of surgical procedure completion times and complications.

The most convenient and frequently employed predictive model is a concave increasing function mapping computation time to expected solution utility. While general meta-deliberation is computationally intractable, this form of profile admits optimal greedy time-slicing methods (Boddy & Dean 1994; Etzioni 1991; Zilberstein & Russell 1996; Horvitz 2001) that may be applied in time-critical situations (on-line). Unfortunately, as we discuss, not all problem-solving procedures are accurately modeled with concave increasing functions.

Meta-deliberation solely based on functions of expected performance ignores the uncertainty of applying problem-solving procedures to new problems. Put another way, problem-solving procedures may exhibit considerable variance in performance across challenges (problem instances). There is a risk that statistical profiles will poorly capture run-time behavior. Correspondingly, narrow variance has been cited as an important property in designing effective problem-solving algorithms (Zilberstein 1996; Boddy & Dean 1994). By contrast, meta-deliberation based on modern portfolio theory (Gomes & Selman 2001) embraces and exploits variance as a tool in improving resource allocation. Uncertainty may also arise in the agent's model of which challenge might occur in the future and how much time will be available for time-critical problem-solving.

Our goal in this work is to develop a meta-deliberation method that retains the on-line efficiency of greedy time-slicing but admits application to problems with (1) non-concave performance profiles, (2) variance in problem-solving performance over the target range of computational challenges, and (3) uncertainty in the parameters of current and future challenges. Our APPROXIMATE AND COMPENSATE method adds flexibility to existing greedy meth-

ods by allowing off-line tuning to better meet real problem-solving conditions. First, the agent approximates problem-solving performance and challenge parameters with concave increasing functions. Second, the agent computes a risk-management factor that compensates for the risk inherent in the approximations. The risk-management factor represents a mean-variance tradeoff that may be derived off-line using any available information. The agent then combines the approximate profiles and compensation factor into an on-line greedy time-slicing meta-deliberation procedure.

This method represents a novel combination of portfolio optimization, meta-deliberation, and continual computation techniques. APPROXIMATE AND COMPENSATE provides an agent the ability to “hedge bets about performance” (Horvitz & Zilberstein 2001) and manage uncertainty.

We first introduce a general terminology for capturing problem-solving performance in a variety of environments. We then present the on-line portion of the APPROXIMATE AND COMPENSATE method and show that the procedure is theoretically sound. We then provide experimental results to explore the effective use of the flexibility provided by APPROXIMATE AND COMPENSATE. We show that off-line tuning of compensation parameters can provide the benefits of fully on-line methods without a significant increase in on-line costs.

MODELS OF META-DELIBERATION

In single-period *continual computation*, Horvitz (Horvitz 2001) differentiates two consecutive time intervals: before and after a discrete change in the state of knowledge of a problem-solving agent. In the first time interval the agent does not know the next computational challenge but may have probabilistic information about which challenges are more likely to occur next and when they are likely to occur. The second time interval is precipitated by some event at which the agent learns the next challenge deterministically. The end of the second time interval occurs when the agent has computed its best response to the known challenge, given the cost of delayed response. The first time interval is referred to as “idle time.” A continual computation agent makes use of this otherwise wasted idle time in preparing to respond to challenges. Following Horvitz (Horvitz 2001), we refer to the first time interval as *precomputation* time. We refer to the second time interval as *reaction* time.

We make the simplifying assumption (Horvitz 2001; Boddy & Dean 1994) of a one-to-one mapping between computational challenges and problem-solving procedures. We focus on the meta-deliberation problem of allocating resources among independent problem-solving procedures for the next unknown challenge. Each procedure is allocated a fraction of the available computation time. Extending our methods to sequential challenges and communicating problem-solving procedures is discussed at the end of this paper.

By allocating computation before the next challenge is known we are precomputing a solution to a challenge that may or may not be presented to us. Thus, the value of allocating computation time to a problem-solving procedure

is only stochastically known in the precomputation time interval. Precomputation time allocated to challenge $i \in I$ is denoted $t_{p,i}$. The time spent reacting to challenge i after the observation point is denoted $t_{r,i}$. The total time allocated to reasoning about challenge i is therefore $t_{p,i} + t_{r,i}$. Although the result of problem-solving depends only on the total time allocation, the *utility* of the result depends on how the total time is split between precomputation and reaction.

Following Horvitz (Horvitz 2001), we define a *value-of-precomputation* function, $f_i(t_{p,i})$, to represent the net utility achieved if challenge i occurs, given precomputation time $t_{p,i}$, and assuming that the agent reasons for an “optimal” length of time once the uncertainty is resolved and the challenge has occurred. What constitutes optimal reactive reasoning depends on the computational model of the agent and the time-critical nature of the environment. In (Parkes & Greenwald 1999) we show how to model the value-of-precomputation function for many common meta-deliberation problems. Table 1 summarizes these results.

A value-of-precomputation function combines a model of the intrinsic value of the problem-solving result with the cost of not responding immediately at the point a challenge is observed. For example, assume the agent deliberates with a run-to-completion (also known as all-or-nothing) algorithm that achieves a value $Value_i$ for challenge i if allocated total time at least $t_{RTC,i}$, and zero value otherwise. Under a soft deadline model, the utility of the response is derived by subtracting the *cost-of-delay*, $Cost_i(t_{r,i})$, from the achieved value ($Cost_i(t) = 0, t \leq 0$). This is a common method for accounting for cost-of-delay (Dean & Wellman 1991). Note that, in this model, cost-of-delay is not incurred until the challenge is observed (i.e. it is only a function of the reaction time). For these problems the agent requires $t_{RTC,i} - t_{p,i}$ time to complete the computation before responding. The value-of-precomputation function for each challenge i is then

$$f_i(t_{p,i}) = Value_i - Cost_i(t_{RTC,i} - t_{p,i}) \quad (1)$$

As a second example, consider an agent that deliberates with an anytime/flexible problem-solving procedure. For this type of computation the intrinsic value of responding to challenge i is captured by an *expected performance profile*, $Value_i(t_{p,i} + t_{r,i})$, that varies as a function of the sum of precomputation and reaction time. Under this model of computation the agent has the flexibility to choose a level of reactive computation that optimizes utility, under a given cost-of-delay model. We can determine an optimal level of reactive computation for any level of precomputation (denoted $t_{r,i}^*(t_{p,i})$) and encode it directly in the value-of-precomputation function, as follows:

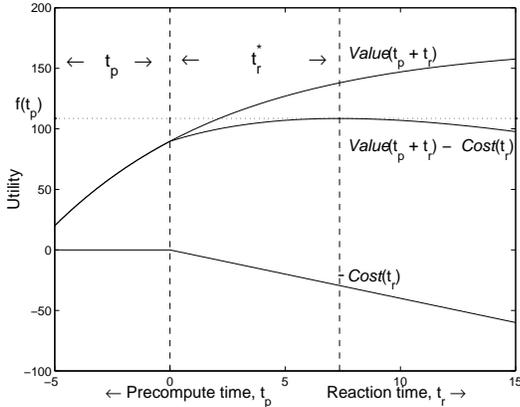
$$f_i(t_{p,i}) = \max_{t_{r,i}} \{ Value_i(t_{p,i} + t_{r,i}) - Cost_i(t_{r,i}) \} \quad (2)$$

Figure 1 illustrates the process of deriving a value-of-precomputation function. The properties of both value function and cost-of-delay combine to determine the shape of the value-of-precomputation function and, as we discuss, the optimality of a greedy time-slicing approach to meta-deliberation. Table 1 indicates the shapes of various combinations of value and cost functions. Commonly found

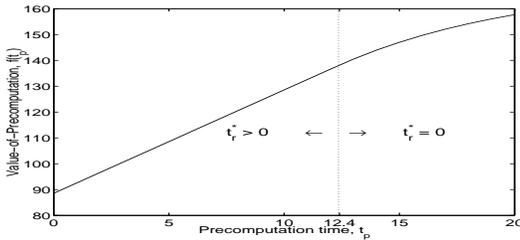
Time-critical Model		Computation Model			
		Run-to completion	Anytime		
			Concave -increasing	Linear -increasing	Convex -increasing
Hard Deadline		×	✓	×	
Cost-of-Delay Function	Convex-increasing	✓	✓	×	
	Linear-increasing	✓	✓	×	
	Concave-increasing	×	×	×	

Table 1: Characterizing the value-of-precomputation functions of common meta-deliberation problems. Meta-deliberation problems optimally solved by greedy time-slicing are marked ✓.

value functions are concave increasing and intuitively capture problem-solving procedures with diminishing returns over time. Convex increasing cost-of-delay functions represent challenges for which it becomes increasingly costly to delay action over time.



(a)



(b)

Figure 1: Determining the value-of-precomputation for a challenge with a linear cost-of-delay function $Cost(t) = 4t$ and an anytime algorithm with a concave increasing performance profile, $Value(t) = 170 - 150e^{-t/8}$: (a) computing the optimal reaction time when allocated precomputation time $t_p = 5$; (b) the complete value-of-precomputation curve, $f_i(t_p)$, $0 \leq t_p \leq 20$. Note that for $t_p > 12.4$ it is optimal to respond immediately upon observing the challenge; while for $t_p \leq 12.4$ it is optimal to spend some additional time problem-solving after the challenge is observed.

An agent meta-deliberates about how to allocate the total precomputation time across possible next challenges, I , $\mathbf{t}_p = (t_{p,1}, \dots, t_{p,I})$. Let $\Pr\{i|E\}$ (abbreviated p_i) be the probability that, given existing knowledge E , challenge $i \in I$ will be the next challenge observed. Meta-deliberation

that maximizes expected value-of-precomputation solves

$$\max_{\mathbf{t}_p} \sum_i p_i f_i(t_{p,i}) \quad (3)$$

such that for each i , $t_{p,i} \geq 0$ and $\sum_i t_{p,i} \leq T$ for some fixed idle time T . It is straightforward to generalize this optimization to distributions over idle time (see Equation 4).

APPROXIMATE AND COMPENSATE

Greedy methods (Boddy & Dean 1994; Etzioni 1991; Zilberstein & Russell 1996; Horvitz 2001) provide effective on-line meta-deliberation in time-critical environments. Unfortunately, the optimality of these methods is limited to a subset of all interesting meta-deliberation problems. In this section we extend the effective use of greedy on-line methods to problems with (1) non-concave performance profiles, (2) variance in problem-solving performance over the target range of computational challenges, and (3) uncertainty in the parameters of current and future challenges. Our method adds flexibility to existing greedy methods by allowing off-line tuning to better meet real problem-solving conditions. Our method mixes the greedy time-slicing method suggested by Horvitz (Horvitz 2001) with a mean-variance technique used in traditional portfolio theory (Markowitz 1959).

In this section we introduce the APPROXIMATE AND COMPENSATE meta-deliberation method. The name of the method is derived from the two ways it extends greedy time-slicing. First, we explicitly acknowledge that oftentimes meta-deliberation must operate on performance profiles that only approximate the “true” performance of problem-solving procedures. Second, we compensate for these approximations with a risk-management factor that represents a mean-variance tradeoff that may be derived off-line using any available information.

More specifically, APPROXIMATE AND COMPENSATE operates on approximate value-of-precomputation functions \hat{f}_i and a risk-aversion parameter γ . The risk-aversion parameter γ is intended to encode either a risk-preference in the meta-deliberation solution when $\hat{f}_i = f_i$, or compensate for using approximations when $\hat{f}_i \neq f_i$. Intuitively, γ captures the agent’s confidence that the approximate value functions will capture run-time performance. If the agent is confident then it will allocate resources as if the approximations are accurate. On the other hand, if the agent is less confident then it will “hedge its bets” by allocating resources to mini-

mize the risk of ending up with a poor solution. We demonstrate experimentally that this is also an effective approximation technique when the true value function is known but non-concave.

Global: \hat{f}_i // value-of-precomputation functions
 p_i // challenge probabilities
 Procedure APPROXIMATE AND COMPENSATE ($\gamma, \Delta, \mathbf{t}_p$) {
 $\mu \leftarrow \sum_i p_i \hat{f}_i(t_{p,i});$
 for $i = 1$ to I {
 $grad_i \leftarrow p_i(\hat{f}_i(t_{p,i} + \Delta) - \hat{f}_i(t_{p,i}));$ }
 // Challenge that maximizes increase in expected value
 $mean \leftarrow \arg_i \max\{grad_i\};$
 // Challenge that maximizes reduction in variance
 $var \leftarrow \arg_i \max\{grad_i(\mu - \hat{f}_i(t_{p,i}))\};$
 // Allocate precomputation time
 $t_{p,mean} \leftarrow t_{p,mean} + (1 - \gamma)\Delta;$
 $t_{p,var} \leftarrow t_{p,var} + \gamma\Delta;$ }

Figure 2: APPROXIMATE AND COMPENSATE is called for each time slice of size Δ to determine how to allocate precomputation time across challenges I according to the mean-variance trade-off encoded in γ . The meta-deliberation solution is incrementally constructed in vector \mathbf{t}_p (where component $t_{p,i}$ is the allocation to challenge i).

The APPROXIMATE AND COMPENSATE method is depicted in Figure 2. In deciding how to allocate a given time slice, APPROXIMATE AND COMPENSATE computes the incremental change in expected value of each possible challenge if allocated that time slice. It then identifies the challenge that would have the maximum effect on increasing total expected value and the challenge that would have the maximum effect on decreasing total variance. The method allocates a fraction of the time slice to the problem-solving procedure for the expected value maximizing challenge and a fraction of the time slice to the problem-solving procedure for the variance minimizing challenge. Allocation fractions are determined by the adjustable risk-aversion parameter γ . As discussed in the next section, when $\gamma = 0$ the method reduces to existing greedy time-slicing methods (Horvitz 2001) that maximize expected value. When $\gamma = 1$ the method greedily minimizes risk. For any $0 < \gamma < 1$ the method selects a tradeoff between mean maximizing and variance minimizing behavior. We later provide an experimental demonstration that, under certain conditions, we can find any solution on the efficient frontier in approximate function space.

SOUNDNESS

The soundness of APPROXIMATE AND COMPENSATE can be demonstrated by its well-behaved properties when $\gamma = 0$ and $\gamma = 1$. The behavior of this method for $0 < \gamma < 1$ is demonstrated experimentally in the next section.

The following theorem states that APPROXIMATE AND COMPENSATE can be used directly to find an optimal solution to the continual computation problem when the agent's

true utility is characterized by concave increasing functions. This solution is found when we set $\hat{f}_i = f_i$ and $\gamma = 0$.

Theorem 1. (Mean-Optimality) APPROXIMATE AND COMPENSATE maximizes expected-value when $\gamma = 0$ for (weakly) concave increasing functions \hat{f}_i as $\Delta \rightarrow 0$.

Proof follows directly from results for continual computation (Horvitz 2001). The greedy strategy is globally optimal because allocating deliberation to a procedure for a locally suboptimal challenge: (a) cannot increase future expected value-of-precomputation from further deliberation on that challenge (weak concavity); (b) has no effect on future expected value-of-precomputation for any other challenges.

The following theorem states that this result holds for all idle times and distributions of challenges. This is possible because the allocation of precomputation time to each challenge in the optimal solution for idle time T is monotonically increasing in T . The meta-deliberation method APPROXIMATE AND COMPENSATE is able to achieve as good a performance as a procedure with distributional information about idle time and unlimited computational resources.

Theorem 2. (On-line Optimality) APPROXIMATE AND COMPENSATE is an optimal on-line procedure for solving

$$\max_{\mathbf{t}_p} \sum_j \Pr(T = T_j) \sum_i p_i \hat{f}_i(t_{p,i}) \quad (4)$$

for any distribution $\Pr(T = T_j)$ over idle times T_j , for (weakly) concave increasing functions \hat{f}_i .

We now show that when \hat{f}_i is a concave increasing function and $\gamma = 1$, APPROXIMATE AND COMPENSATE will locally minimize variance in value-of-precomputation, given the approximation, from challenge to challenge.

Theorem 3. (Local Variance-Optimality) When $\gamma = 1$, APPROXIMATE AND COMPENSATE allocates precomputation in the current time-slice to the challenge that minimizes the variance in approximate value-of-precomputation at the end of the slice, for (weakly) concave increasing approximation functions \hat{f}_i .

Proof. Let $\text{Var}(\mathbf{t}_p) = \sum_i p_i (\hat{f}_i(t_{p,i}))^2 - (\sum_i p_i \hat{f}_i(t_{p,i}))^2$, denote the variance of approximate value-of-precomputation from challenge to challenge, given an allocation \mathbf{t}_p . The allocation over the next time slice that achieves the greatest local decrease in variance, is to the challenge that minimizes $\partial \text{Var}(\mathbf{t}_p) / \partial t_{p,i} = 2p_i(\partial \hat{f}_i(t_{p,i}) / \partial t_{p,i})(\hat{f}_i(t_{p,i}) - \mu)$, where $\mu = \sum_i p_i \hat{f}_i(t_{p,i})$. As $\Delta \rightarrow 0$, APPROXIMATE AND COMPENSATE allocates the next time-slice to precomputation on this challenge. Note that there must always be at least one challenge that decreases variance while the variance is non-zero because there must be a challenge with $\hat{f}_i(t_{p,i}) < \mu$, and $\partial \hat{f}_i(t_{p,i}) / \partial t_{p,i}$ is non-decreasing for concave increasing functions \hat{f}_i . \square

We can also show that APPROXIMATE AND COMPENSATE is an on-line globally optimal variance minimizing meta-deliberation method for the special case when all challenges are (1) equally likely, and (2) have linear approximate value-of-precomputation functions with the same gradient, $\hat{f}_i(t_{p,i}) = \alpha t_{p,i}$, for some constant α .

EXPERIMENTAL RESULTS

In this section we experimentally demonstrate the effective use of APPROXIMATE AND COMPENSATE with $0 < \gamma < 1$. In the first class of problems, the agent does not have an exact value-of-precomputation function for each challenge but, rather, a distribution over possible functions. This may be due to an incomplete model of the environment, or approximations introduced in the computational model. In the second class of problems, the agent is faced with non-concave value-of-precomputation functions and chooses a concave approximation that enables efficient on-line meta-deliberation, and compensates for that approximation through the selection of an optimal risk-aversion parameter off-line. Finally, we provide a risk interpretation of the conditions under which an agent would choose $\gamma \neq 0$ off-line.

PROTOTYPE FUNCTIONS

In the following experiments we consider an agent facing two possible challenges. The first challenge is modeled with an exact value-of-precomputation function, while the uncertainty in performance of the problem-solving procedure for the second challenge is modeled by a parameterized distribution of value-of-precomputation functions. The response to each challenge is compute by run-to-completion problem-solving procedures with $t_{RTC,1} = t_{RTC,2} = 10$. Challenge 1 occurs with probability $p_1 = 0.6$, and has an exact cost-of-delay profile $Cost(t_{r,1}) = t_{r,1}^{1.6}$. Challenge 2 occurs with probability $p_2 = 0.4$, and has a uncertain cost-of-delay profile that is distributed according to a parameterized distribution, $Cost(t_{r,2}) = t_{r,2}^x$, where x is distributed uniformly between 1.4 and 2.0 (i.e. $x \sim U(1.4, 2.0)$). The value of an immediate response to both challenges is $Value_1 = Value_2 = 30$. The value-of-precomputation functions are computed as $f_i(t_{p,i}) = Value_i - Cost_i(t_{RTC,i} - t_{p,i})$ (Equation 1). The distribution of value-of-precomputation functions for challenge 2 that is induced by the distribution over cost-of-delay functions is shown in Figure 3.

We assume the agent does not know before the challenge occurs how much idle time is available for continual computation. In the next section we discuss the effects of uncertain idle times on the use of completely off-line stochastic optimization methods. In this section we focus primarily on on-line methods, with or without off-line tuning.

We consider five alternative techniques for solving the continual computation problem stated above. The first technique (1) is to perform off-line stochastic optimization using the full distribution over value-of-precomputation functions, assuming prior knowledge of idle time. This technique is used to provide an upper bound on performance. The other four techniques use some form of greedy time-slicing over an approximate model of value-of-precomputation for challenge 2, selecting (2) a simple linear approximation and $\gamma = 0$, (3) a simple linear approximation and γ optimized off-line, (4) a prototype approximation drawn from the distribution of functions and $\gamma = 0$, and (5) a prototype approximation optimally derived off-line from the full distribution and $\gamma = 0$. The results of these experiments are summarized

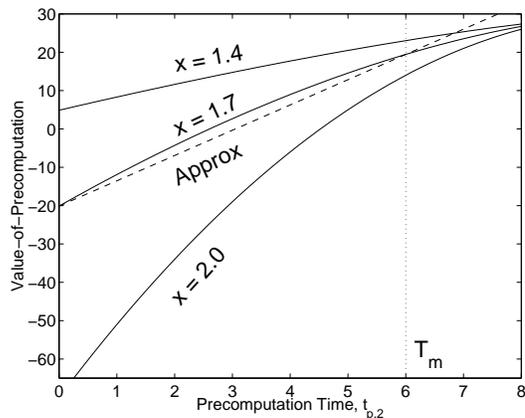


Figure 3: Value-of-precomputation $f_2(t_{p,2})$ for an incomplete meta-deliberation model, represented as a parameterized distribution: $f_2(t_{p,2}) = 30 - (10 - t_{p,2})^x$, where $x \sim U(1.4, 2.0)$, for $t_{p,2} \leq 10$. A linear approximation to the median curve ($x = 1.7$) is fitted to cross at $t_{p,2} = 0$ and $t_{p,2} = 6$.

in Table 2.

Techniques (2-4) use APPROXIMATE AND COMPENSATE with simple prototypes, with only experiment (3) tuning γ off-line. The linear approximation used in techniques (2) and (3) is chosen to meet the median curve of the distribution at $t_{p,2} = 0$ and $t_{p,2} = T_m$, where T_m is the mean idle time. For example, the linear approximation for $T \sim U(4, 8)$ is shown in Figure 3. The off-line optimization of γ in technique (3) is implemented through exhaustive search, with the value of γ assessed with stochastic sampling and simulation. Technique (4) uses the median curve from the distribution. Technique (5) uses APPROXIMATE AND COMPENSATE with an optimal prototype. The optimal in-distribution prototype is computed off-line with the following stochastic optimization: exhaustively search over the space of curves, assessing the performance of each curve by simulating APPROXIMATE AND COMPENSATE with sampling from the distribution over idle times and actual value-of-precomputation functions.

Meta-deliberation procedure	γ	Expected value
(1) Optimal with hindsight	-	6.27
(2) Linear Approximation	0	4.08
(3) Linear Approximation	0.46	4.71
(4) Simple Prototype ($x = 1.7$)	$\gamma = 0$	4.61
(5) Optimal Prototype ($x = 1.75$)	$\gamma = 0$	4.71

Table 2: Average expected value-of-precomputation for a distribution of idle times and an incomplete model of meta-deliberation.

Table 2 shows that, for this simple example, we achieve as good a performance with a simple approximation (linear in this case) and off-line optimization of γ , as with off-line selection of an optimal prototype ($x = 1.75$) for $\gamma = 0$. In this table the results are averaged over idle times uniformly distributed between 4 and 8. Figure 4 shows the results over each idle time.

Comparing curves (a) and (c) we see that greedy time-slicing with $\gamma = 0$ and a simple linear approximation (curve (a)) is biphase – initially allocating deliberation to challenge 1, and then to challenge 2. The benefit of our mean-variance approach is seen in curve (c), in which off-line optimization of γ enables a not-so-greedy on-line meta-deliberation strategy. In this strategy APPROXIMATE AND COMPENSATE allocates more deliberation to challenge 2 from the start of precomputation, reducing approximate-risk, and improving performance over $\gamma = 0$.

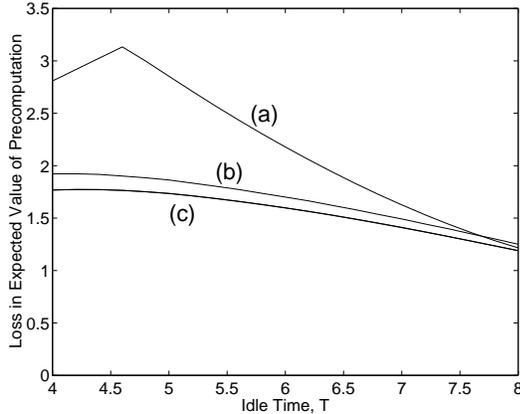


Figure 4: Loss in expected value-of-precomputation versus idle time, T , with respect to the best performance with hindsight, for: (a) Linear off-line approximation, $\gamma = 0$; (b) Simple prototype ($x = 1.7$), $\gamma = 0$; (c) Linear off-line approximation and compensation, $\gamma = 0.46$ and Optimal prototype ($x = 1.75$), $\gamma = 0$.

An additional property of optimizing γ for a simple approximation is that, while the search space over optimal prototypes grows with the number of challenges with uncertain performance models, the computation of γ^* remains tractable. Furthermore, in this simple example, APPROXIMATE AND COMPENSATE outperforms the best “reasonable” guess at a prototype (the median prototype), and perform as well as the optimal prototype.

APPROXIMATING NON-CONCAVE FUNCTIONS

In the following experiments we demonstrate the effectiveness of APPROXIMATE AND COMPENSATE on hard meta-deliberation problems for which greedy time-slicing is not provably optimal. Consider again two challenges, each with run-to-completion algorithms, but now with *non-concave* cost-of-delay functions. Challenge 1 occurs with probability $p_1 = 0.55$, and has run-to-completion time $t_{RTC,1} = 15$. Challenge 2 occurs with probability $p_2 = 0.45$, and has run-to-completion time $t_{RTC,2} = 20$. The value of an immediate response to either challenge is $Value_1 = Value_2 = 2$, and each challenge has a sigmoidal cost-of-delay, $0 \leq Cost_i(t_{r,i}) = e^{-(t_{r,i}/10)^3} < 1$. This cost function models the case in which an initial delay is not too costly, and once a response is significantly delayed, a further delay is unimportant. The value-of-precomputation function for each challenge is shown in Figure 5. We assume that the idle time is

uniformly distributed, $T \sim U(T_m - T_u, T_m + T_u)$, with mean T_m and uncertainty T_u . Note that, in contrast to the optimality results for concave curves, meta-deliberation with non-concave curves is sensitive to idle time.

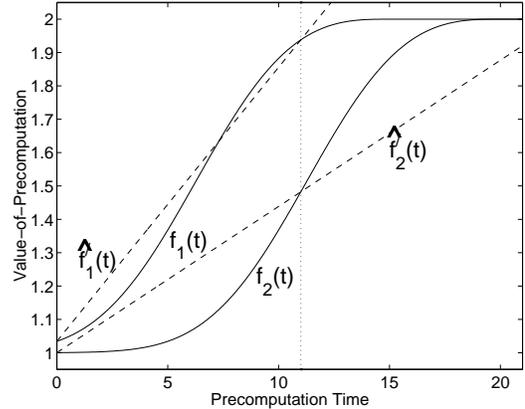


Figure 5: Value-of-precomputation versus precomputation time for challenges with soft-deadlines and run-to-completion algorithms. $Value = 2$, $Cost(t) = 1 - e^{-(t/10)^3}$, $t_{RTC,1} = 15$, $t_{RTC,2} = 20$. Linear approximations fitted at $t_p = 0$ and $t_p = 11$.

We consider four alternative techniques for solving the continual computation problem stated above. The first technique (1) is to perform off-line stochastic optimization using the full distribution over value-of-precomputation functions, assuming prior knowledge of idle time. This technique is used to provide an upper bound on performance. The next two techniques use simple linear approximations, \hat{f}_i , to the non-concave functions and perform greedy time-slicing with (2) $\gamma = 0$, and (3) γ^* optimally tuned off-line to a given model of idle time uncertainty. The fourth technique (4) proceeds greedily on the true value-of-precomputation curves, f_i . The simple linear approximations used in techniques (2-3) are fit to the true curves at $t_p = 0$ and $t_p = T_m/2$. Two such approximations for $T_m = 22$ are shown in Figure 5. Technique (3) uses exhaustive search and sampling and simulation of APPROXIMATE AND COMPENSATE to select an optimal risk-aversion parameter.

The results of these experiments are summarized in Table 3. These results depict each technique with idle time uncertainty $T_u = 3$ and mean drawn from $T_m \in \{10, 16, 22, 30\}$.

Meta-deliberation procedure	Mean idle time, T_m			
	10	16	22	30
(1) Optimal	1.47	1.56	1.76	1.98
(2) Linear ($\gamma = 0$)	1.47	1.56	1.61	1.93
(3) Linear (γ^*)	1.47	1.55	1.75	1.97
($\leftrightarrow \gamma^*$)	(0.0)	(0.0)	(0.59)	(0.78)
(4) Exact ($\gamma = 0$)	1.47	1.55	1.62	1.94

Table 3: Average expected value-of-precomputation for idle times $T \sim U(T_m - 3, T_m + 3)$, for $T_m \in \{10, 16, 22, 30\}$.

The performance of APPROXIMATE AND COMPENSATE

with off-line tuning of γ^* (technique (3)) is competitive with the best solution with hindsight (technique (1)) for all idle time distributions. The performance of greedy meta-deliberation on the true value-of-precomputation curves (technique (4)) is suboptimal for $T_m \in \{22, 30\}$. All the approaches perform well for small idle times, $T_m \in \{10, 16\}$. However, for large idle times, greedy time-slicing cannot take advantage of prior knowledge that the slope of the true value-of-precomputation functions changes direction.

In Figure 6 we depict the performance of APPROXIMATE AND COMPENSATE if we choose a single compensation factor (i.e. constant portfolio) off-line without knowledge of the mean idle time. Performance of techniques (2-4) are shown relative to the optimal technique (1). The choice of $\gamma = 0.59$ optimizes the online performance for idle times $T \sim U(19, 25)$.

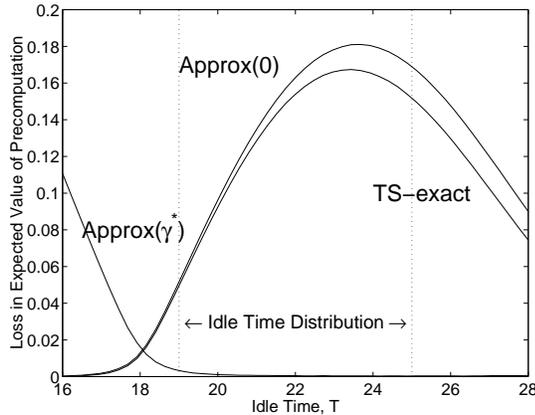


Figure 6: Comparison of loss due to on-line meta-deliberation versus optimal technique (1) for technique (2) (Approx(0)), technique (3) (Approx($\gamma^* = .59$)), and technique (4) (TS-exact).

Finally, we consider the effect of uncertainty in idle time, T_u , on the performance of APPROXIMATE AND COMPENSATE. Figure 7 shows the performance for $T_m = 22$, for idle time uncertainty $T_u \in [0, 13]$. The performance with an optimal off-line risk-aversion parameter (technique (3)) decreases as uncertainty increases, finally performing slightly worse than time-slicing on the exact curves (technique (4)). The methodology works well when the distributional information about the likely idle time can be used off-line to improve performance. This is not possible for large idle time uncertainty and non-concave value-of-precomputation curves because a meta-deliberation allocation that is optimal for small idle times may not also be optimal for longer idle times. In this case there is only a marginal advantage in having access to distributional information on idle time.

EFFICIENT FRONTIER

Mean-variance analysis is a decision-analytic method in modern portfolio theory for choosing an optimal investment portfolio (Markowitz 1959). Mean-variance analysis is performed over the “efficient frontier”, a set of portfolios that dominate all other portfolios. The expected utility of a port-

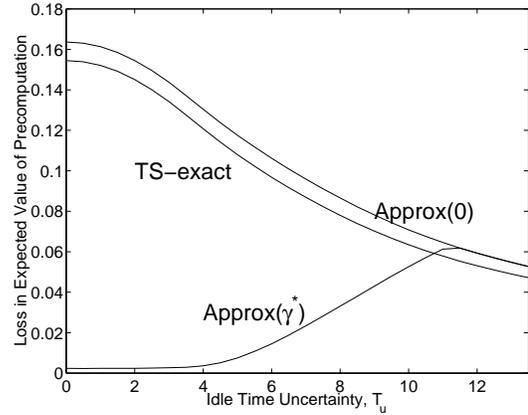


Figure 7: Average expected value-of-precomputation for idle times $T \sim U(22 - T_u, 22 + T_u)$ versus idle time uncertainty, T_u . Comparison of loss due to on-line meta-deliberation versus optimal technique (1) for technique (2) (Approx(0)), technique (3) (Approx($\gamma^* = .59$)), and technique (4) (TS-exact).

folio is assumed to increase with expected single-period return, but decrease with variance in period-to-period return (risk) – and a portfolio is efficient if it achieves a greater expected return than any other portfolio with the same risk.

In Figure 8 we demonstrate that APPROXIMATE AND COMPENSATE given (weakly) concave increasing approximate value-of-precomputation functions and any $\gamma \in [0, 1]$ generates solutions that lie on the efficient frontier. Thus, this method may be used to choose solutions other than that which maximizes expected value-of-precomputation. This gives a risk interpretation of the conditions under which an agent would choose $\gamma \neq 0$ off-line.

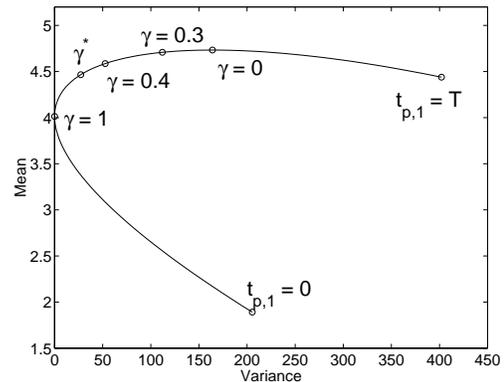


Figure 8: Variance versus mean for deliberation portfolio selection in the approximate \hat{f} value-of-precomputation space for the problem shown in Figure 3. The efficient frontier lies between $\gamma = 0$ and $\gamma = 1$, and $\gamma^* = 0.46$ represents optimal off-line risk-compensation.

We are currently investigating if APPROXIMATE AND COMPENSATE always finds *only* solutions on the efficient frontier and whether or not it can find *all* points on the efficient frontier. These properties would enable off-line opti-

mization that is more efficient than searching the complete space of allocations.

DISCUSSION

We develop a meta-deliberation method that retains the on-line efficiency of greedy time-slicing but admits application to problems with (1) non-concave performance profiles, (2) variance in problem-solving performance over the target range of computational challenges, and (3) uncertainty in the parameters of current and future challenges. Our APPROXIMATE AND COMPENSATE method adds flexibility to existing greedy methods through approximation and off-line tuning. We show that the method is sound and experimentally demonstrate how to use off-line tuning to provide the benefits of fully on-line methods without a significant increase in on-line costs. Important future work is to provide more guidance in how to choose good approximation functions off-line.

Boddy and Dean (1994) describe how the difference in variance of problem-solving procedures with similar expected performance can greatly affect meta-deliberation decisions. They discuss quantifying the cost of meta-deliberation with approximate profiles. Zilberstein and Russell (1996) approximate probabilistic profiles by varying the granularity of tabular representations or through closed-form normal distribution approximations. They point out the subsequent errors inherent in closed-form approximations.

Continual computation with two distinct states of knowledge may be generalized to problems in which changes to this problem-solving “belief state” occur in increments over time. We are additionally exploring extensions of APPROXIMATE AND COMPENSATE to sequential probabilistic challenges. When evaluating meta-deliberation over sequences of challenges we must consider both competition for shared resources across time (Boddy & Dean 1994) and dependencies between current deliberations and future challenges.

In our experimental results APPROXIMATE AND COMPENSATE is shown to perform well but, not as well as off-line optimization with hindsight. We can show that this baseline for performance can not be attained for many problems. Intuitively, an optimal allocation assuming idle time $T + 1$ can not necessarily be reached by starting with an optimal allocation for idle time T . Thus, without prior information about idle time, we cannot produce an idle time allocation that is optimal for both T and $T + 1$. Put another way, idle time allocation to each challenge must be monotonically non-decreasing. We can pose optimization problems that take these additional constraints into account. For example, we might consider the class of *constant-portfolio* strategies that allocate the same proportion of idle time to each challenge for all idle times.

Acknowledgements This research is sponsored in part by a Veteran’s Health Administration Medical Informatics fellowship, in part by a Drexel/MCP-Hahnemann Research Synergies award, in part by a National Science Foundation (NSF) Instrumentation Award under grant CISE-9986105, in part by the National Library of Medicine under grant R01 LM05764, and in part by NSF grant SBR 97-08965.

References

- Boddy, M., and Dean, T. 1994. Decision-theoretic deliberation scheduling for problem solving in time-constrained environments. *Artificial Intelligence* 67(2):245–286.
- Clarke, J. R.; Trooskin, S. Z.; Doshi, P. J.; Greenwald, L. G.; and Mode, C. J. 2001. Time to laparotomy for intra-abdominal bleeding from trauma does affect survival for delays up to 90 minutes. *Journal of the American Association of Trauma Surgeons*.
- Dean, T., and Wellman, M. P. 1991. *Planning and Control*. Morgan Kaufmann.
- Etzioni, O. 1991. Embedding decision-analytic control in a learning architecture. *Artificial Intelligence* 49(1-3):129–160.
- Garvey, A., and Lesser, V. 1993. Design-to-time real-time scheduling. *IEEE Transactions on Systems, Man, and Cybernetics* 23(6):1491–1502.
- Georgeff, M. P., and Lansky, A. L. 1987. Reactive reasoning and planning. In *Proceedings AAAI-87*.
- Gomes, C. P., and Selman, B. 2001. Algorithm portfolios. *Artificial Intelligence Special Issue on Tradeoffs under Bounded Resources* 126(1-2):43–62.
- Greenwald, L., and Dean, T. 1994. Solving time-critical decision-making problems with predictable computational demands. In *Second International Conference on AI Planning Systems*, 25–30.
- Horvitz, E. J., and Breese, J. S. 1990. Ideal partition of resources for metareasoning. Technical Report KSL-90-26, Knowledge Systems Laboratory, Stanford University.
- Horvitz, E., and Lengyel, J. 1997. Perception, attention, and resources: A decision-theoretic approach to graphics rendering. In Geiger, D., and Shenoy, P. P., eds., *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence (UAI-97)*, 238–249. San Francisco: Morgan Kaufmann Publishers.
- Horvitz, E., and Zilberstein, S. 2001. Computational tradeoffs under bounded resources. *Artificial Intelligence Special Issue on Tradeoffs under Bounded Resources* 126(1-2):1–4.
- Horvitz, E. J.; Suermondt, H. J.; and Cooper, G. F. 1989. Bounded conditioning: Flexible inference for decisions under scarce resources. In *UW89*, 182–193.
- Horvitz, E. 2001. Principles and applications of continual computation. *Artificial Intelligence Special Issue on Tradeoffs under Bounded Resources* 126(1-2):159–196.
- Markowitz, H. M. 1959. *Portfolio Selection*. Wiley, New York.
- Parkes, D. C., and Greenwald, L. G. 1999. On the optimality of greedy meta-deliberation. Technical report, Drexel University, Department of Mathematics and Computer Science.
- Zilberstein, S., and Russell, S. J. 1993. Anytime sensing, planning and action: A practical model for robot control. In *Proc. 13th International Joint Conference on Artificial Intelligence (IJCAI-93)*, 1402–1407.
- Zilberstein, S., and Russell, S. 1996. Optimal composition of real-time systems. *Artificial Intelligence* 82:181–213.
- Zilberstein, S. 1993. *Operational Rationality through Compilation of Anytime Algorithms*. Ph.D. Dissertation, University of California at Berkeley.
- Zilberstein, S. 1996. Using anytime algorithms in intelligent systems. *AI Magazine* 73–83.