Solving Stochastic Nonlinear Resource Allocation Problems Using a Hierarchy of Twofold Resource Allocation Automata

Ole-Christoffer Granmo and B. John Oommen, *Fellow*, *IEEE*

Abstract—In a multitude of real-world situations, resources must be allocated based on incomplete and noisy information. However, in many cases, incomplete and noisy information render traditional resource allocation techniques ineffective. The decentralized Learning Automata Knapsack Game (LAKG) was recently proposed for solving one such class of problems, namely the class of *Stochastic Nonlinear Fractional Knapsack Problems*. Empirically, the LAKG was shown to yield a superior performance when compared to methods which are based on traditional parameter estimation schemes. This paper presents a completely new online Learning Automata (LA) system, namely the *Hierarchy of Twofold Resource Allocation Automata* (H-TRAA). In terms of contributions, we first of all, note that the primitive component of the H-TRAA is a *Twofold Resource Allocation Automaton* (TRAA) which possesses novelty in the field of LA. Second, the paper contains a formal analysis of the TRAA, including a rigorous proof for its convergence. Third, the paper proves the convergence of the H-TRAA itself. Finally, we demonstrate empirically that the H-TRAA provides *orders of magnitude* faster convergence compared to the LAKG for simulated data pertaining to two-material unit-value functions. Indeed, in contrast to the LAKG, the H-TRAA scales sublinearly. Consequently, we believe that the H-TRAA opens avenues for handling demanding real-world applications such as the allocation of sampling resources in large-scale web accessibility assessment problems. We are currently working on applying the H-TRAA solution to the web-polling and sample-size detection problems applicable to the world wide web.

Index Terms—Nonlinear knapsack problems, hierarchical learning, learning automata, stochastic optimization, resource allocation.

1 Introduction

1.1 Motivation

In a multitude of real-world situations, resources are often to be allocated based on incomplete and noisy information. Such resource allocation problems are particularly intriguing because, in many cases, incomplete and noisy information render traditional optimization techniques ineffective. In this paper, we address one such model which can be translated into a family of problems:

> *Imagine that you have to allocate a limited amount of time among n different activities. The problem is such that spending a time instant on an activity randomly produces one of two possible outcomes—the time instant is either spent "fruitfully" or "unfruitfully." In this generic setting, your goal is to maximize the expected amount of fruitfully spent time. Unfortunately, you are only given the following information regarding the activities:*
>
> 1. *each instant of time spent on an activity has a certain probability of being fruitful, and*
> 2. *this probability decreases with the amount of time spent on the activity.*
>
> *To render the problem even more realistic, you do not have access to the probabilities themselves. Instead, you must rely on solving the problem by means of trial-and-failure, i.e., by attempting different allocations and observing the resulting random outcomes.*

To pose the problem in a practical perspective, we present below two simple applications of the problem in time management and in monitoring webpages.

Example 1—Time Management. Students frequently encounter a real-life instantiation of the above intriguing problem. A student that pursues several different topics in a semester has to decide how to allocate his working hours among the topics. After a day of work, the student will have some idea of how much he has learned during the day, allowing him to assess his current allocation of working hours. Rather than over specializing in a single topic and treating the other topics superficially, seeking overall mastery of the topics could be a wise choice in this situation. However, the amount of time required to master a topic will vary, simply because the nature of a topic influences the student's "learning curve" for that specific topic. Thus, finding an optimal allocation in this problem must involve "trial and error," and unknown success probabilities, as in our present model.

Example 2—Webpage Monitoring. The problem also appears in other settings. The world wide web is, for instance, an extremely vast resource-thirsty field, which probably consumes a major portion of the computing resources available today. Searching, updating, and examining webpages is, undoubtedly, one of the primary tasks done by both individuals and companies today. This, in turn, leads to numerous extremely interesting real-life resource allocation and scheduling problems,

- *O.-C. Granmo is with the Department of Information and Communication Technology (ICT), Faculty of Engineering and Science, University of Agder, Service Box 422, NO-4604 Kristiansand, Norway. E-mail: ole.granmo@uia.no.*
- *B.J. Oommen is with the School of Computer Science, Carleton University, Ottawa, ON K1S 5B6, Canada and the University of Agder in Grimstad, Norway. E-mail: oommen@scs.carleton.ca.*

Manuscript received 2 July 2007; revised 15 Aug. 2009; accepted 22 Sept. 2009; published online 28 Dec. 2009.
Recommended for acceptance by F. Lombardi.
For information on obtaining reprints of this article, please send e-mail to: tc@computer.org, and reference IEEECS Log Number TC-2007-07-0291.
Digital Object Identifier no. 10.1109/TC.2009.189.

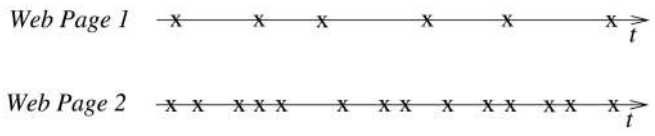

Fig. 1. Webpage changes occurring over time. An "x" on the time lines denotes that the respective webpage has changed. Observe that the occurrence of this event is not observable to the outside world unless the webpage is polled.

and the problem studied here is pertinent to one such problem, the so-called "web-polling" problem.

Webpage monitoring consists of repeatedly polling a selection of webpages so that the user can detect changes that occur over time. Clearly, as this task can be prohibitively expensive, in practical applications, the system imposes a constraint on the *maximum* number of webpages that can be polled per time unit. This bound is dictated by the governing communication bandwidth, and by the speed limitations associated with the processing. Since only a fraction of the webpages can be polled within a given unit of time, the problem which the system's analyst encounters is one of determining which webpages are to be polled. In such cases, a reasonable choice of action is to choose webpages in a manner that maximizes the number of changes detected, and the optimal allocation of the resources again involves "trial and error." As illustrated in Fig. 1, webpages may change with varying frequencies (that are unknown to the decision maker), and changes appear more or less randomly.

Furthermore, as argued elsewhere, [7], [8], the probability that an individual webpage poll uncovers a change on its own decreases monotonically with the polling frequency used for that webpage. Consider that time is discrete, with the time interval length T to be the atomic unit of decision making. In each time interval, every single webpage i has a constant probability u_i of remaining *unchanged*. Furthermore, when a webpage is updated/changed, the update is available for detection only until the webpage is updated again. After that, the original update is considered lost. For instance, each time a newspaper webpage is updated, previous news items are replaced by the most recent ones. By way of example, consider the scenario that a webpage remains unchanged in any single time step with probability 0.5. Then, polling the webpage uncovers new information with probability $1 - 0.5^3 = 0.875$ if the webpage is polled every third time step (i.e., with frequency $\frac{1}{3}$) and $1 - 0.5^2 = 0.75$ if the webpage is polled every second time step. As seen, increasing the polling frequency reduces the probability of discovering new information on each polling.

The problem that we study here has direct applications to this web-polling problem, and to the problem of determining the optimal sample size required for estimation purposes—both of which are addressed here briefly, but in more detail elsewhere [9].

1.2 Formal Problem Formulation

The above problem instances can be formulated as *Stochastic Nonlinear Fractional Equality Knapsack (NEFK) Problems* as exemplified earlier [7], [8], [9]. Such a formulation permits an analytically rigorous treatment of the problem.

In order to appreciate the qualities of the Stochastic NEFK Problem, it is beneficial to view the problem in light of the classical *linear* Fractional Knapsack (FK) Problem. Indeed, the Stochastic NEFK Problem generalizes the latter problem in two significant ways. Both of the two problems are *briefly* defined below.

The linear fractional knapsack (FK) problem. The linear FK problem is a classical continuous optimization problem which also has applications within the field of resource allocation. The problem involves n materials of different value v_i per unit volume, $1 \leq i \leq n$, where each material is available in a certain amount $x_i \leq b_i$. Let $f_i(x_i)$ denote the value of the amount x_i of material i, i.e., $f_i(x_i) = v_i x_i$. The problem is to fill a knapsack of fixed volume c with the material mix $\vec{x} = [x_1, \ldots, x_n]$ of maximal value $\sum_1^n f_i(x_i)$ [2].

The nonlinear equality FK (NEFK) problem. One important extension of the above classical problem is the *Nonlinear Equality* FK problem with a separable and concave objective function. The problem can be stated as follows [13]:

$$\begin{aligned} &\text{maximize} \quad f(\vec{x}) = \sum_1^n f_i(x_i), \\ &\text{subject to} \quad \sum_1^n x_i = c \quad \text{and} \quad \forall i \in \{1, \ldots, n\}, x_i \geq 0. \end{aligned}$$

Since the objective function is considered to be concave, the value function $f_i(x_i)$ of each material is also concave. This means that the derivatives of the material value functions $f_i(x_i)$ with respect to x_i (hereafter denoted f_i'), are nonincreasing. In other words, the material value *per unit volume* is no longer constant as in the linear case, but decreases with the material amount, and so the optimization problem becomes:

$$\begin{aligned} &\text{maximize} \quad f(\vec{x}) = \sum_1^n f_i(x_i), \text{ where } f_i(x_i) = \int_0^{x_i} f_i'(x_i) dx_i, \\ &\text{subject to} \quad \sum_1^n x_i = c \quad \text{and} \quad \forall i \in \{1, \ldots, n\}, x_i \geq 0. \end{aligned}$$

Efficient solutions to the latter problem, based on the principle of Lagrange multipliers, have been devised. In short, the optimal value occurs when the derivatives f_i' of the material value functions are equal, subject to the knapsack constraints [3], [5]:

$$\begin{aligned} &f_1'(x_1) = \cdots = f_n'(x_n), \\ &\sum_1^n x_i = c \quad \text{and} \quad \forall i \in \{1, \ldots, n\}, x_i \geq 0. \end{aligned}$$

The stochastic NEFK problem. In this paper, we generalize the above nonlinear equality knapsack problem. First of all, we let the material value per unit volume for any x_i be a *probability* function $p_i(x_i)$. Furthermore, we consider the distribution of $p_i(x_i)$ to be *unknown*. That is, each time an amount x_i of material i is placed in the knapsack, we are only allowed to observe an instantiation of $p_i(x_i)$ at x_i, and not $p_i(x_i)$ itself. Given this stochastic environment, we intend to devise an online incremental scheme that learns the mix of materials of maximal *expected* value, through a series of informed guesses. Thus, to clarify issues, we are provided with a knapsack of fixed volume c, which is to be filled with a mix of n different materials. However, unlike the NEFK, in the Stochastic NEFK Problem the unit volume value of a material i, $1 \leq i \leq n$, is a random quantity—it takes the value 1 with probability $p_i(x_i)$ and the value 0 with probability $1 - p_i(x_i)$, respectively. As an additional complication, $p_i(x_i)$ is nonlinear in the sense that it decreases monotonically with x_i, i.e., $x_{i_1} \leq x_{i_2} \Leftrightarrow p_i(x_{i_1}) \geq p_i(x_{i_2})$.

Since unit volume values are random, we operate with expected unit volume values rather than the actual unit volume values. With this understanding, and the above perspective in mind, the expected value of the amount x_i of material i, $1 \leq i \leq n$, becomes $f_i(x_i) = \int_0^{x_i} p_i(u)du$. Accordingly, the expected value per unit volume[1] of material i becomes $f_i'(x_i) = p_i(x_i)$. In this stochastic and nonlinear version of the FK problem, the goal is to fill the knapsack so that the expected value $f(\vec{x}) = \sum_1^n f_i(x_i)$ of the material mix contained in the knapsack is maximized. Thus, we aim to:

$$\begin{aligned} \text{maximize} \quad & f(\vec{x}) = \sum_1^n f_i(x_i), \text{ where } f_i(x_i) = \int_0^{x_i} p_i(u)du, \\ & \qquad \text{and } p_i(x_i) = f_i'(x_i), \\ \text{subject to} \quad & \sum_1^n x_i = c \quad \text{and} \quad \forall i \in \{1, \ldots, n\}, x_i \geq 0. \end{aligned}$$

A fascinating property of the above problem is that the amount of information available to the decision maker is limited—the decision maker is only allowed to observe the current unit value of each material (either 0 or 1). That is, each time a material mix is placed in the knapsack, the unit value of each material is provided to the decision maker. The actual outcome probabilities $p_i(x_i), 1 \leq i \leq n$, however, remain *unknown*. As a result of the latter, the expected value of the material mix must be maximized by means of trial-and-failure, i.e., by experimenting with different material mixes and by observing the resulting random unit-value outcomes.

In the interest of posing our paper within the larger body of scientific research, we present here, in all brevity, some of the more recent studies on the stochastic knapsack problem. The problem, in its virgin form, has been recently studied by Gibson et al. [6], who proposed an intelligent agent-based stochastic ruler approach for it. The interesting feature of this result is that these authors have examined the scenario in which the system has to execute a sequence of resource allocation decisions over time, but the availability of the indivisible resources at future epochs is uncertain due to the actions of the competitors. By utilizing a multiperiod bounded multiple-choice knapsack framework, these authors proposed a general discrete stochastic optimization model that solved the corresponding knapsack problem with a nonlinear objective function. Similarly, from a purely theoretical perspective, Lin et al. [14] studied the stochastic knapsack problem when the system permitted switch-over policies and dynamic pricing. A new approximate solution which uses the benefits of adaptivity was recently proposed by Dean et al. [4]. This result is fascinating because the authors show that adaptivity provides only a constant-factor improvement, and this is done by demonstrating a greedy nonadaptive algorithm that approximates the optimal adaptive policy within a factor of 7. Dean et al. [4] have also designed an adaptive polynomial-time algorithm which approximates the optimal adaptive policy within a factor of $5 + \epsilon$, for any constant $\epsilon > 0$.

From a more practical standpoint, Perry and Hartman [20] have modeled the multiperiod, single resource capacity reservation problem as a dynamic, stochastic, multiple knapsack problem, where the state space grows exponentially with the number of knapsacks. The solution they proposed led to a fast approximate solution inasmuch as determining the optimal solution is computationally intractable. In the same vein, Sachs [22] enforced a stochastic knapsack model to solve the capacity evaluation problem for multiradio access networks. The interesting feature of his solution is that the model was able to consider the nonuniform geographic distribution of both the radio link capacity and the traffic load. The stochastic knapsack formulation has also been used to distributing layered encoded videos by caching [12].

1.3 State of the Art

To the best of our knowledge, prior to our work reported in [9], our targeted stochastic NEFK problem was not addressed in the literature before. However, several studies on related problems have been reported. For example, the works of Dean et al. [4] and Steinberg and Parks [24] consider solution policies for stochastic generalizations of the so-called NP-hard *linear* integer knapsack problem. In these papers, value distributions were considered known and constant, making dynamic programming a viable solution. Another variant of the knapsack problem is found in [21] where a deterministic knapsack is used, however, with objects arriving to and departing from the knapsack at random times. The optimization problem considered was to accept/block arriving objects so that the average value of the knapsack content is maximized.

The first reported generic treatment of the stochastic NEFK problem itself can be found in [9]. Various instantiations of the problem have, however, appeared sporadically, particularly within the web monitoring domain. In these latter instantiations, the unknown parameters are *estimated* by means of a tracking phase where webpages are polled mainly for estimation purposes [19], [27]. One major disadvantage of such an approach is that the parameter estimation phase significantly delays the implementation of an optimal solution. This disadvantage is further aggravated in *dynamic* environments where the optimal solution changes over time, introducing the need for parameter reestimation [7].

In contrast to the above approaches, we base our work on the principles of Learning Automata (LA) [15], [25]. LA have been used to model biological systems [26] and have attracted considerable interest in the last decade because they can learn the optimal actions when operating in (or interacting with) unknown stochastic environments. Furthermore, they combine rapid and accurate convergence with low computational complexity.

The novel Learning Automata Knapsack Game (LAKG) scheme that we proposed in [9] does not rely on estimating parameters and can be used to solve the stochastic NEFK problem in both static and dynamic settings. Indeed, empirical results verify that the LAKG finds the optimal solution with arbitrary accuracy, guided by the principle of Lagrange Multipliers. Furthermore, the empirical results show that the performance of the LAKG is superior to that of parameter-estimation-based schemes, both in static and dynamic environments. Accordingly, we believe that the LAKG can be considered to represent the state of the art when it concerns research on the stochastic NEFK problem. This landmark is now extended to develop the Twofold Resource Allocation Automaton (TRAA) (which, in itself is the first reported LA which is *artificially* ergodic[2]), and its hierarchical version, the H-TRAA.

1. We, hereafter, use $f_i'(x_i)$ to denote the derivative of the expected value function $f_i(x_i)$ with respect to x_i.

2. LA which have been artificially made *absorbing* to yield specific properties, have been earlier reported [16]. However, we are not aware of any LA which, in essence are absorbing, but which have been made artificially *ergodic*.

1.4 Contributions of This Paper

The contributions of this paper are the following:

1. We report the first *analytical* results for schemes that solve the Stochastic NEFK Problem.
2. We propose a novel scheme for the *two-material* resource allocation problem, namely the *TRAA*. As mentioned, from the perspective of LA, the TRAA, in itself, is the first reported LA which is *artificially* rendered ergodic.
3. We prove that the TRAA is asymptotically optimal.
4. We report the first *hierarchical* solution to the Stochastic NEFK Problem, based on a hierarchy of TRAAs, namely the H-TRAA, which is also proven to be asymptotically optimal.
5. We verify empirically that the H-TRAA provides orders of magnitude faster convergence compared to the LAKG.

As a result of the above contributions, we believe that the H-TRAA is a viable realistic strategy for solving demanding real-world problems such as the optimal allocation of sampling resources in large-scale web accessibility assessment [23] and other problems related to the world wide web [9].

1.5 Paper Organization

The paper is organized as follows: In Section 2, we present the TRAA for the *two-material* problem, and prove its asymptotic optimality. We then propose how TRAAs can be arranged in a hierarchy for solving *multimaterial* Stochastic NEFK Problems. We proceed in Sections 3 and 4 to verify empirically that the H-TRAA provides orders of magnitude faster convergence compared to the LAKG when applied to two specific detection functions, and to a specific application domain, respectively. Indeed, we shall present results that clearly demonstrate that the H-TRAA allows us to tackle 32,768-material problems in *real time*. Finally, we offer suggestions for further work (including those to the two application domains alluded to earlier) before we conclude the paper in Section 5.

2 A Hierarchy of Twofold Resource Allocation Automata (H-TRAA)

2.1 Overview of the H-TRAA Solution

In order to put our work in the right perspective, we start this section by providing a brief review of the concepts and the solution found in [9], which are also relevant for more "primitive" variants of the knapsack problem.

As indicated in Section 1, solving the classical linear FK problem involves finding the most valuable mix $\vec{x}^* = [x_1^*, \ldots, x_n^*]$ of n materials that fits within a knapsack of fixed capacity c. The material value per unit volume for each material i is given as a constant v_i, and each material is available in a certain amount $x_i \leq b_i$, $1 \leq i \leq n$. Accordingly, the value of the amount x_i of material i, $f_i(x_i) = v_i x_i$, is linear with respect to x_i. In other words, the derivative of $f_i(x_i)$—i.e., the material value per unit volume—is fixed: $f_i'(x_i) = v_i$. Because a fraction of each material can be placed in the knapsack, the following greedy algorithm from [2] finds the most valuable mix: *Take as much as possible of the material that is most valuable per unit volume. If there is still room, take as much as possible of the next most valuable material. Continue until the knapsack is full.*

Let us now generalize this and assume that the material unit volume values are *random* variables with *constant* and *known* distributions. Furthermore, for the sake of conceptual clarity, let us only consider binary variables that *either* instantiate to the values of 0 or 1. Since the unit volume values are random, let p_i denote the probability of the unit volume value $v_i = 1$ for material i, $1 \leq i \leq n$, which means that the probability of the unit volume value $v_i = 0$ becomes $1 - p_i$. With some insight, it becomes evident that under such conditions, the above greedy strategy can again be used to maximize the *expected* value of the knapsack, simply by selecting material based on the *expected* unit volume values, $E[v_i] = 0 \times (1 - p_i) + 1 \times p_i$, rather than actual unit volume values.

The above indicated solution is, of course, inadequate when the p_is are unknown. Furthermore, the problem becomes even more challenging when the p_is are no longer constant, but rather depend on their respective material amounts x_i, $1 \leq i \leq n$. Let $p_i(x_i)$ denote the probability that the current unit volume value of material i is $v_i = 1$, given that the amount x_i has already been placed in the knapsack. Then, the expected value per unit volume of material i, $1 \leq i \leq n$, becomes $E[v_i] = 0 \times [1 - p_i(x_i)] + 1 \times p_i(x_i) = p_i(x_i)$, and accordingly, the expected value of the amount x_i becomes $f_i(x_i) = \int_0^{x_i} p_i(u)du$.

Our aim, then, is to find a scheme that moves toward optimizing the following NEFK problem online:

$$\begin{aligned} \text{maximize} \quad & f(\vec{x}) = \sum_1^n f_i(x_i), \text{ where } f_i(x_i) = \int_0^{x_i} p_i(u)du, \\ & \qquad \text{and } p_i(x_i) = f_i'(x_i), \\ \text{subject to} \quad & \sum_1^n x_i = c \quad \text{and} \quad \forall i \in \{1, \ldots, n\}, x_i \geq 0. \end{aligned}$$

Note that we allow only instantiations of the material values per unit volume to be observed. That is, each time an amount x_i of material i is placed in the knapsack, an instantiation v_i at x_i is observed.

Because of the above intricacies, we approach the problem by relying on informed material mix *guesses*, i.e., by experimenting with different material mixes and learning from the resulting random unit volume value outcomes. We shall assume that x_i is any number in the interval $(0, 1)$. The question of generalizing this will be considered later. The crucial issue that we have to address, then, is that of determining how to change our current guesses on x_i, $1 \leq i \leq n$. We shall attempt to do this in a discretized manner by subdividing the unit interval into N points $\{\frac{1}{N+1}, \frac{2}{N+1}, \ldots, \frac{N}{N+1}\}$, where N is the resolution of the learning scheme. We will see that a larger value of N will ultimately imply a more accurate solution to the knapsack problem.

At this juncture, it is pertinent to mention that although the rationale for this updating is the stochastic point location solution proposed by Oommen [17], the two schemes are quite distinct for the following reasons:

1. The method proposed in [17] assumes the existence of an Oracle which informs the LA whether to go "right" or "left." In our application domain, this now has to be *inferred* by the system.

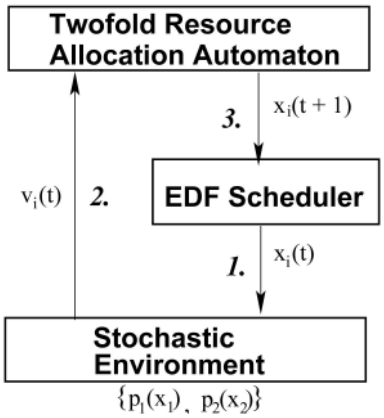

Fig. 2. The TRAA interacting with a Scheduler and an unknown Stochastic Environment.

2. The method proposed in [17] assumes that there is only a single LA in the picture. Here, we specifically understand that there are multiple LAs organized in a hierarchy—each of them being constrained to work together with the others.[3]
3. In [17], the problem of analyzing scenarios with space varying responses from the environment was left open. This problem is tackled in the present paper.
4. As opposed to the scheme in [17], our present approach is also applicable to dynamic (time varying) environments.
5. There is a "huge" fundamental difference between the LA which we devise here and the work of [17]. Unlike the latter, in which the system is truly ergodic, our present LA would be *absorbing* if the end-states of the probability space are also included. However, to forcefully render this present machine ergodic, we have artificially made the LA ergodic by *excluding* these states from the set of possible probability values. This makes the analysis both distinct and quite fascinating. As mentioned earlier, we are not aware of any LA which, in essence are absorbing, but which have been made artificially *ergodic*.

2.2 Details of the TRAA Solution

2.2.1 Design of the TRAA Solution

We first present our LA-based solution to *two-material* Stochastic NEFK Problems. The two-material solution forms a critical part of the hierarchic scheme for multiple materials that is presented subsequently. As illustrated in Fig. 2, our solution to two-material problems constitutes of three modules:

1. a Stochastic Environment,
2. the TRAA itself, and
3. an Earliest Deadline First (EDF) Scheduler.

We first detail each of the three modules, before we analyze the overall *feedback connection* between them. Finally, we prove that the TRAA that we have developed in this section is asymptotically optimal for two-material Stochastic NEFK Problems.

3. It is conceivable that this problem can be resolved with a single LA possessing an extended number of actions. But we do not recommend it for scalability reasons—the action space would grow exponentially.

Stochastic Environment. The *Stochastic Environment* for the two-material case can be characterized by:

1. the capacity c of the knapsack; and
2. two-material unit volume value probability functions $p_1(x_1)$ and $p_2(x_2)$.

In brief, if the amount x_i of material i is suggested to the Stochastic Environment, the Environment replies with a unit volume value $v_i = 1$ with probability $p_i(x_i)$ and a unit volume value $v_i = 0$ with probability $1 - p_i(x_i)$, $i \in \{1, 2\}$. It should be emphasized that to render the problem both interesting and nontrivial, we assume that $p_i(x_i)$ is unknown to the TRAA.

TRAA. The scheme which attempts to learn the optimal allocation $\vec{x}^* = [x_1^*, x_2^*]$ can be described as follows: A finite fixed structure automaton with the states $s(t) \in \{1, 2, \ldots, N\}$ is used to decide the allocation of resources among the two materials. Let the current state of the automaton be $s(t)$. Furthermore, let $q_{s(t)}$ refer to the fraction $\frac{s(t)}{N+1}$, and let $r_{s(t)}$ refer to the fraction: $1 - q_{s(t)}$. Then, the automaton's current guess is $\vec{x} = [q_{s(t)}, r_{s(t)}]$.

If the Stochastic Environment tells the automaton that the unit volume value of material i is $v_i(t)$ at time t, the automaton updates its state as follows:

$$s(t+1) := s(t) + 1 \quad \textbf{If } \textbf{rand}() \leq r_{s(t)} \textbf{ and } v_i(t) = 1 \textbf{ and } 1 \leq s_i(t) < N \textbf{ and } i = 1, \tag{1}$$

$$s(t+1) := s(t) - 1 \quad \textbf{If } \textbf{rand}() \leq q_{s(t)} \textbf{ and } v_i(t) = 1 \textbf{ and } 1 < s_i(t) \leq N \textbf{ and } i = 2, \tag{2}$$

$$s(t+1) := s(t) \qquad \textbf{Otherwise}. \tag{3}$$

Fig. 3 shows the resulting stochastic transition graphs for resolution $N = 5$. The upper graph shows the transitions for feedback from the Stochastic Environment on material 1, and the graph below shows the transitions for feedback on material 2. Notice how the stochastic state transitions are designed to offset the learning bias introduced by accessing the materials with frequencies proportional to $\vec{x} = [q_{s(t)}, r_{s(t)}]$. Also observe that the overall learning scheme does not produce any absorbing states, and is, accordingly, ergodic supporting dynamic environments. The effect of these properties is analyzed in the next section.

Finally, after the automaton has had the opportunity to change its state, it provides output to the EDF Scheduler. That is, it outputs the material amounts $\vec{x} = [q_{s(t+1)}, r_{s(t+1)}]$ that have been changed.

EDF Scheduler. The Scheduler takes material amounts $\vec{x} = [x_1, \ldots, x_n]$ as its input (for the two-material case the input is $\vec{x} = [x_1, x_2]$). The purpose of the Scheduler is:

1. to provide accesses to the Stochastic Environment in a sequential manner, and
2. to make sure that the unit volume value functions are accessed with frequencies proportional to $\vec{x}$.

The reader should note that our scheme does not rely on accessing the unit volume value functions sequentially with frequencies proportional to $\vec{x}$ for solving the knapsack problem. However, this restriction is obviously essential for

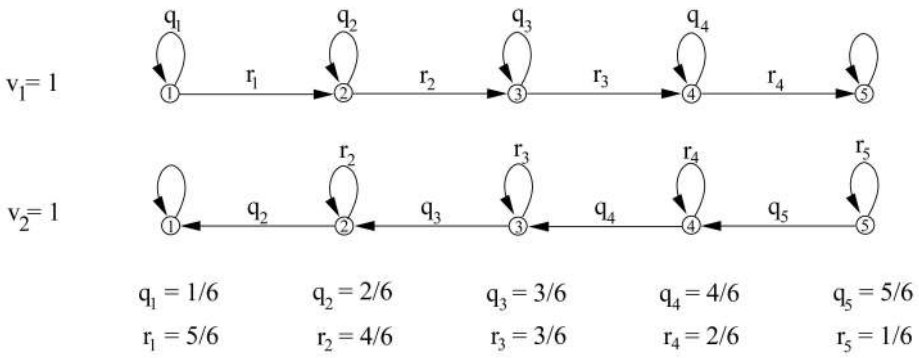

Fig. 3. The stochastic transition graphs of a TRAA with resolution $N = 5$.

solving the problem *incrementally* and *online* (or rather in a "real-time" manner). Note that since it, in some cases, may be essential to access each unit volume value function with a constant period and not randomly (for example, in the earlier-alluded-to problem which analyzes webpage polling), we use the EDF Scheduling to access the functions according to $\vec{x}$.

2.2.2 Analysis of the TRAA Solution

In this section, we characterize the optimal solution to a Stochastic NEFK Problem. Thereafter, we analyze the feedback connection of the TRAA and the Stochastic Environment—we prove that the TRAA is asymptotically optimal in the sense that it can find material allocations arbitrarily close to the solution of the Stochastic NEFK Problem.

Lemma 1. *The material mix $\vec{x} = [x_1, \ldots, x_n]$ is a solution to a given Stochastic NEFK Problem if 1) the derivatives of the expected material amount values are all equal at $\vec{x}$, 2) the mix fills the knapsack, and 3) every material amount is positive, i.e.:*

$$f_1'(x_1) = \cdots = f_n'(x_n),$$
$$\sum_1^n x_i = c \text{ and } \forall i \in \{1, \ldots, n\}, x_i \geq 0.$$

The above lemma is based on the well-known principle of Lagrange Multipliers [3], [5], and its proof is, therefore, omitted here for the sake of brevity. Instead, we will start by analyzing the *two-material* problem and the TRAA. Multiple TRAAs will then be organized in a hierarchy with the aim of tackling *n-material* problems.

For the two-material problem, let $\vec{x}^* = [x_1^*, x_2^*]$ denote a solution, as defined above. Note that since x_2^* can be obtained from x_1^*, we will concentrate on finding x_1^*.

Theorem 1. *The TRAA solution scheme specified by 1)-3) is asymptotically optimal.*

Proof. Our aim is to prove that as the resolution, N, is increased indefinitely, the expected value of the TRAA output, $x_1(t)$, converges toward the solution of the problem, x_1^*, implying that:

$$\lim_{N\to\infty} \lim_{t\to\infty} E[x_1(t)] \to x_1^*.$$

We shall prove the above by analyzing the properties of the underlying Markov chain, which is specified by the EDF Scheduler, the rules 1)-3) (the TRAA), and the Environment. As can be seen from 1)-3), the states of the chain are the integers $j \in \{1, 2, \ldots, N\}$. In brief, rules 1)-3), when interacting with the EDF Scheduler and the Environment, obey the Markov chain with transition matrix $H = [h_{ij}]$, where

$$h_{j,j-1} = r_j \cdot p_2(r_j) \cdot q_j, \quad 1 < j \leq N, \tag{4}$$

$$h_{j,j+1} = q_j \cdot p_1(q_j) \cdot r_j, \quad 1 \leq j < N, \tag{5}$$

$$h_{j,j} = 1 - h_{j,j-1} - h_{j,j+1}, \quad 1 < j < N, \tag{6}$$

and, accordingly,

$$h_{1,1} = 1 - h_{1,2}, \tag{7}$$

$$h_{N,N} = 1 - h_{N,N-1}. \tag{8}$$

Clearly, H represents a single closed communicating class whose periodicity is unity. The chain is ergodic, and the limiting probability vector is given by the eigenvector of H^T corresponding to eigenvalue unity. Let this vector be $\Pi = [\pi_1, \pi_2, \ldots, \pi_N]$. Then, Π satisfies:

$$\begin{bmatrix} h_{1,1} & h_{1,2} & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ h_{2,1} & h_{2,2} & h_{2,3} & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & h_{3,2} & h_{3,3} & h_{3,4} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & h_{N-2,N-3} & h_{N-2,N-2} & h_{N-2,N-1} & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & h_{N-1,N-2} & h_{N-1,N-1} & h_{N-1,N} \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & h_{N,N-1} & h_{N,N} \end{bmatrix}^T$$
$$\begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \cdot \\ \cdot \\ \pi_{N-2} \\ \pi_{N-1} \\ \pi_N \end{bmatrix} = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \cdot \\ \cdot \\ \pi_{N-2} \\ \pi_{N-1} \\ \pi_N \end{bmatrix}. \tag{9}$$

The details of solving (9) are quite cumbersome, and we undertake it now. Observe that our aim is to prove that the probability mass of Π lies arbitrarily close to the solution of the knapsack problem, $\vec{x}^* = [x_1^*, x_2^*]$, as N goes to infinity. Before we go through the fine details, we outline the proof strategy as follows: We first explicitly solve for the quantities $\{\pi_i\}$ by solving the underlying difference equations. We then define a function U that forms an upper bound for Π. We proceed to show that the upper bound goes to zero outside an arbitrarily close vicinity of x_1^*, as the resolution, N, goes to infinity. Accordingly, since Π is a probability distribution, and since U is its upper bound, increasing the resolution toward infinity, moves the probability mass of Π arbitrarily close to x_1^*.

The details of the proof follow. Our first step is to reformulate the individual row-wise equations from the matrix (9) recursively. Expanding the first row of (9) yields:

$$\pi_1 \cdot h_{1,1} + \pi_2 \cdot h_{2,1} = \pi_1 \Rightarrow \pi_2 = \frac{(1 - h_{1,1}) \cdot \pi_1}{h_{2,1}} = \frac{h_{1,2}}{h_{2,1}} \cdot \pi_1. \tag{10}$$

Expanding the second row of (9) and substituting (10) yields:

$$\pi_1 \cdot h_{1,2} + \pi_2 \cdot h_{2,2} + \pi_3 \cdot h_{3,2} = \pi_2 \Rightarrow \pi_3 = \frac{h_{2,3}}{h_{3,2}} \cdot \pi_2. \tag{11}$$

Arguing in a similar way in a row-by-row manner, it can be seen[4] that

$$\pi_{k-1} = \frac{h_{k,k-1}}{h_{k-1,k}} \cdot \pi_k, \tag{12}$$

for $0 < k \leq N$, which, on reversing the recursion, yields for $0 \leq k < N$,

$$\pi_{k+1} = \frac{h_{k,k+1}}{h_{k+1,k}} \cdot \pi_k. \tag{13}$$

Let

$$\alpha(x_1, N) = \left\lfloor \frac{x_1}{\frac{1}{N+1}} \right\rfloor \text{ and } \beta(x_1, N) = \left\lceil \frac{x_1}{\frac{1}{N+1}} \right\rceil.$$

Clearly, $[\frac{\alpha(x_1,N)}{N+1}, \frac{\beta(x_1,N)}{N+1}]$ is the interval that most accurately approximate x_1 given the resolution N. In particular, with $z = \alpha(x_1^*, N)$, the solution x_1^* is found in the interval $[\frac{z}{N+1}, \frac{z+1}{N+1}]$. The crucial part of our proof is to reformulate Π in terms of π_z and π_{z+1}, using (12) and (13). More specifically, for $j \in \{1, \ldots, z-1\}$ we have:

$$\pi_j = \pi_z \cdot \prod_{k=z}^{j+1} \frac{h_{k,k-1}}{h_{k-1,k}}. \tag{14}$$

Correspondingly, and arguing in an analogous manner, for $j \in \{z+2, \ldots, N\}$ we have:

$$\pi_j = \pi_{z+1} \cdot \prod_{k=z+1}^{j-1} \frac{h_{k,k+1}}{h_{k+1,k}}. \tag{15}$$

4. We omit the laborious algebraic steps in the interest of readability.

In other words, we represent Π in terms of two of its components: π_z and π_{z+1}.

We are now ready to define the upper bound U for Π:

$$U[i, z] = \begin{cases} \pi_z \cdot M^{z-i}, & \text{if } i \leq z, \\ \pi_{z+1} \cdot M^{i-(z+1)}, & \text{if } i \geq z+1, \end{cases} \tag{16}$$

where:

$$M = \max\left[\max_{k \leq z}\left\{\frac{h_{k,k-1}}{h_{k-1,k}}\right\}, \max_{k \geq z+1}\left\{\frac{h_{k,k+1}}{h_{k+1,k}}\right\}\right]. \tag{17}$$

As seen, the definition of M clearly makes U an upper bound for Π.

Our final goal is to show that as the resolution N goes to infinity, U goes to zero outside an arbitrarily close vicinity of x_1^*:

$$\lim_{N \to \infty} U[\alpha(x_1, N), \alpha(x_1^*, N)] \to 0, \quad \text{if } x_1 \neq x_1^*. \tag{18}$$

We shall argue that the latter is guaranteed to happen if we have

$$0 < \frac{h_{k,k-1}}{h_{k-1,k}} < 1 \quad \text{for} \quad k \in \{2, \ldots, z\}$$

and $0 < \frac{h_{k,k+1}}{h_{k+1,k}} < 1$ for $k \in \{z+1, \ldots, N-1\}$, because then we get $0 < M < 1$. We argue this by considering the equilibrium (asymptotic) value of $E[\pi(t)]$ for any finite N. This argument can be separated into three different cases as in [17]:

1. The first case is when $\frac{z}{N+1}$ is close to zero. In this case, the maximum is quickly reached and then geometrically falls away.
2. When $\frac{z}{N+1}$ is close to 1, the value of π_i geometrically increases but when the maximum is reached it quickly falls away. For both these cases when $N \to \infty$, most of the probability mass will be centered in a small interval around z.
3. The third case is slightly more complex because it involves $\frac{z}{N+1}$ being away from either end. This case must be broken down into two distinct geometric series, one representing the geometric series from π_1 to π_z and the other from π_{z+1} to π_N. The first series increases until it reaches the maximum at π_Z. The increase is geometric (or rather, exponential as $N \to \infty$), and the geometric ratio is bounded by the bound given by the quantity M above. The second series starts at the maximum at the value π_{z+1} and then decreases until π_N is reached. Again, the decrease is geometric (i.e., exponential as $N \to \infty$), and the geometric ratio is bounded by the quantity M above. In this case, the probability mass will be centered within a small interval around $\frac{z}{N+1}$ and $\frac{z+1}{N}$ as $N \to \infty$ because of the law of the *sum* of the elements of a geometric series possessing a common ratio which is greater than unity.

First of all, since the difference between $\frac{k}{N+1}$ and $\frac{k-1}{N+1}$ goes to zero as N goes to infinity, and since $p_1(x)$ is continuous, we have:

$$\lim_{N \to \infty} \frac{h_{k,k-1}}{h_{k-1,k}} = \lim_{N \to \infty} \frac{r_k \cdot p_2(r_k) \cdot q_k}{q_k \cdot p_1(q_k) \cdot r_k} \tag{19}$$

$$= \lim_{N\to\infty} \frac{p_2(r_k)}{p_1(q_k)}. \quad (20)$$

Second, from Lemma 1, we can conclude that $p_1(q_k) > p_2(r_k)$ for $k \in \{2, \ldots, z\}$. Therefore, $0 < \frac{h_{k,k-1}}{h_{k-1,k}} < 1$ for $k \in \{2, \ldots, z\}$ as N goes to infinity.

Showing that we have $0 < \frac{h_{k,k+1}}{h_{k+1,k}} < 1$ for $k \in \{z+1, \ldots, N-1\}$ follows analogously, and the proof is left out here for the sake of brevity.

Accordingly, Π must go to zero outside an arbitrarily close vicinity of x_1^* as the resolution N goes to infinity. This, in turn, means that the probability mass of Π will lie arbitrarily close to x_1^*. In other words, the TRAA is asymptotically optimal. □

2.3 Details of the H-TRAA Solution

2.3.1 Design of the H-TRAA Solution

In this section, we propose a hierarchical scheme for solving n-material problems. The scheme takes advantage of the TRAA's ability to solve two-material problems asymptotically, by organizing them hierarchically.

Construction of hierarchy. The hierarchy of TRAAs, which we, hereafter, will refer to as H-TRAA, is constructed as follows[5]: First of all, the hierarchy is organized as a balanced binary tree with depth $D = \log_2(n)$. Each node in the hierarchy can be related to three entities: 1) a set of materials, 2) a partitioning of the material set into two subsets of equal size, and 3) a dedicated TRAA that allocates a given amount of resources among the two subsets.

Root node. The hierarchy root (at depth 1) is assigned the complete set of materials $S_{1,1} = \{1, \ldots, n\}$. These n materials are partitioned into two disjoint and exhaustive subsets of equal size: $S_{2,1}$ and $S_{2,2}$. An associated TRAA, $T_{1,1}$, decides how to divide the full knapsack capacity c (which, for the sake of notational correctness will be referred to as $c_{1,1}$) among the two subsets. That is, subset $S_{2,1}$ receives the capacity $c_{2,1}$ and subset $S_{2,2}$ receives the capacity $c_{2,2}$, with $c_{2,1} + c_{2,2} = c_{1,1}$. Accordingly, *this* TRAA is given the power to prioritize one subset of the materials at the expense of the other.

Nodes at depth d. Node $j \in \{1, \ldots, 2^{d-1}\}$ at depth d (where $1 < d \leq D$) refers to: 1) the material subset $S_{d,j}$, 2) a partitioning of $S_{d,j}$ into the subsets $S_{d+1,2j-1}$ and $S_{d+1,2j}$, and 3) a dedicated TRAA, $T_{d,j}$. Observe that since level $D+1$ of the H-TRAA is nonexistent, we use the convention that $S_{D+1,2j-1}$ and $S_{D+1,2j}$ refer to the primitive materials being processed by the leaf TRAA, $T_{D,j}$. Assume that the materials in $S_{d,j}$ has, as a set, been assigned the capacity $c_{d,j}$. The dedicated TRAA, then, decides how to allocate the assigned capacity $c_{d,j}$ among the subsets $S_{d+1,2j-1}$ and $S_{d+1,2j}$. That is, subset $S_{d+1,2j-1}$ receives the capacity $c_{d+1,2j-1}$ and subset $S_{d+1,2j}$ receives the capacity $c_{d+1,2j}$, with $c_{d+,2j-1} + c_{d+1,2j} = c_{d,j}$.

At depth D, then, each individual material can be separately assigned a fraction of the overall capacity by way of recursion, using the above allocation scheme.

Interaction of H-TRAA with EDF scheduler and environment. As in the single TRAA case, H-TRAA interacts with an EDF Scheduler, which suggests which unit volume value function $p_i(x_i)$ to access next. A response is then generated from the Stochastic Environment using $p_i(x_i)$. This response is given to all the TRAAs that were involved in determining the material amount x_i, that is, the TRAAs in the hierarchy that have allocated capacity to a material subset that contains material i. Finally, a new candidate material mix $\vec{x} = [x_1, \ldots, x_n]$ is suggested by the H-TRAA to the EDF Scheduler.

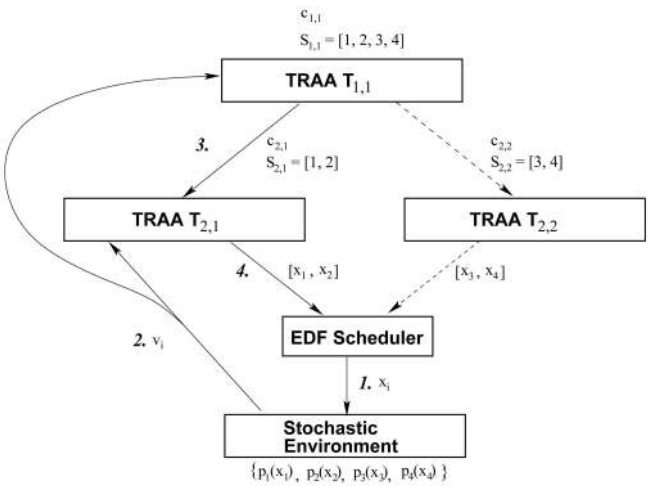

Fig. 4. A H-TRAA interacting with a Scheduler and an unknown Stochastic Environment as explained in Example 1.

Example 1. Consider a four-material problem. Fig. 4 shows the associated hierarchy, constructed as described above. At the root level, the TRAA $T_{1,1}$ divides the knapsack capacity among the two-material subsets $\{1, 2\}$ and $\{3, 4\}$, respectively, related to TRAA $T_{2,1}$ and $T_{2,2}$. At the level below, then, the TRAA $T_{2,1}$ allocates its share of the capacity among material 1 and material 2, while TRAA $T_{2,2}$ assigns *its* share of the capacity to material 3 and material 4. Based on the present assignment at time t, the EDF Scheduler selects material i, suggesting the amount $x_i(t)$ to the Stochastic Environment. The Stochastic Environment, in turn, responds with a randomly drawn material unit volume value, $v_i(t)$, using the probability value function $p_i(x_i)$. By way of example, if $i = 2$, the latter feedback is given to TRAAs $T_{1,1}$ and $T_{2,1}$, which update their states accordingly, and the feedback loop continues.

2.3.2 Analysis of the H-TRAA Solution

In the previous section, we proved that an individual TRAA is asymptotically optimal. We will now consider the H-TRAA and prove its optimality. More specifically, we shall show that if each individual TRAA in the hierarchy has solved its own two-material problem, a solution to the complete n-material Knapsack Problem has also been produced.

Theorem 2. *Let $T_{d,j}$ be an arbitrary TRAA at level d of the H-TRAA associated with the node whose index is j. Then, if every single TRAA, $T_{d,j}$, in the H-TRAA has found a local solution with proportions $c_{d+1,2j-1}$ and $c_{d+1,2j}$ satisfying*

$$f'_{d+1,2j-1}(c_{d+1,2j-1}) = f'_{d+1,2j}(c_{d+1,2j}),$$

5. We assume that $n = 2^\gamma, \gamma \in \mathbb{N}^+$, for the sake of clarity. If the number of materials is less than this, we can assume the existence of additional materials whose values are "zero," and who thus are not able to contribute to the final optimal solution.

the overall Knapsack Problem involving n materials that are hierarchically placed in $\log_2 n$ levels of TRAAs, also attains the global optimum solution.

Proof. We intend to prove the above theorem by means of induction, using the hierarchical H-TRAA structure defined in the paragraph titled *Construction of Hierarchy*.

Basis. The *Basis* case concerns the nodes at the leaves, which, indeed, deal with the primitive materials themselves. Let a and b ($a, b \in \{1, \ldots, n\}$) be any two materials processed by a TRAA, $T_{D,u}$, at a leaf node (i.e., at depth $\mathrm{D} = \log_2 n$) in the H-TRAA. The latter decides how to allocate an assigned capacity $c_{D,u}$ among the two materials a and b, with relative proportions x_a and x_b, respectively. Observe that since a and b are the only two materials relevant to this TRAA, by virtue of the construction of the TRAA, $\frac{x_a}{x_a+x_b}$ and $\frac{x_b}{x_a+x_b}$ are the conditional probabilities of choosing a and b, respectively, conditioned on the event that the knapsack had only to be filled with these primitive materials. Since, by virtue of Theorem 1, we know that the TRAA will find a local solution $[x_a, x_b]$, the foundation of the solution determined by the Lagrangian yields:

$$f'_a(x_a) = f'_b(x_b) \Rightarrow f'_{D+1,2u-1}(c_{D+1,2u-1}) = f'_{D+1,2u}(c_{D+1,2u}), \quad \text{with } c_{D+1,2u-1} + c_{D+1,2u} = c_{D,u},$$

thus proving the basis of the induction.

Induction step. Consider any interior-node TRAA $T_{d,j}$ whose index at depth d is j in the H-TRAA hierarchy. The TRAA associated with this node decides how to allocate an assigned capacity $c_{d,j}$ among two disjoint subsets $S_{d+1,2j-1} = \{\alpha_1, \ldots, \alpha_m\}$ and $S_{d+1,2j} = \{\beta_1, \ldots, \beta_m\}$ of *composite* materials, where each α_i and β_i is, in itself, a primitive material. To simplify notation, let $\vec{\alpha} = \{\alpha_1, \ldots, \alpha_m\}$ and $\vec{\beta} = \{\beta_1, \ldots, \beta_m\}$. Observe that the union of the sets $\vec{\alpha}$ and $\vec{\beta}$ is the input to the present TRAA, and the task of *this* TRAA is to assign the current knapsack capacity, $c_{d,j}$, so as to satisfying the Lagrangian solution for these two mutually exclusive and exhaustive subsets. Let $T_{d,j}$ assign the relative proportions to $\vec{\alpha}$ and $\vec{\beta}$ by the quantities $x_{\vec{\alpha}}$ and $x_{\vec{\beta}}$. Observe that since $\vec{\alpha}$ and $\vec{\beta}$ are the only two materials[6] relevant to this TRAA, by virtue of the construction of the TRAA,

$$\frac{x_{\vec{\alpha}}}{x_{\vec{\alpha}} + x_{\vec{\beta}}} \text{ and } \frac{x_{\vec{\beta}}}{x_{\vec{\alpha}} + x_{\vec{\beta}}}$$

are the conditional probabilities of choosing $\vec{\alpha}$ and $\vec{\beta}$, respectively, conditioned on the event that the knapsack had only to be filled with these composite materials $\vec{\alpha}$ and $\vec{\beta}$. The solution to this TRAA will thus satisfy:

$$f'_{\vec{\alpha}}(x_{\vec{\alpha}}) = f'_{\vec{\beta}}(x_{\vec{\beta}}), \text{ where} \tag{21}$$

$$f'_{\vec{\alpha}}(x_{\vec{\alpha}}) = \sum_{\alpha_i \in \vec{\alpha}} \frac{x_{\alpha_i}}{\sum_{\alpha_j \in \vec{\alpha}} x_{\alpha_j}} f'_{\alpha_i}(x_{\alpha_i}) \text{ and} \tag{22}$$

$$f'_{\vec{\beta}}(x_{\vec{\beta}}) = \sum_{\beta_i \in \vec{\beta}} \frac{x_{\beta_i}}{\sum_{\beta_j \in \vec{\beta}} x_{\beta_j}} f'_{\beta_i}(x_{\beta_i}). \tag{23}$$

6. The fact that these are composite materials is irrelevant to the present TRAA. It merely treats $\vec{\alpha}$ and $\vec{\beta}$ as individual materials.

Since each α_i and β_i is a primitive material, and we are working our way *up* the H-TRAA hierarchy, we can invoke the inductive hypothesis to relate x_{α_i} and x_{β_i} for all i. By virtue of the inductive hypothesis and the Lagrangian solution at *every* level up the H-TRAA till level d, we know that for both of the material subsets $S_{d+1,2j-1}$ and $S_{d+1,2j}$ the following are true:

$$f'_{\alpha_1}(x_{\alpha_1}) = \cdots = f'_{\alpha_m}(x_{\alpha_m}), \tag{24}$$

$$f'_{\beta_1}(x_{\beta_1}) = \cdots = f'_{\beta_m}(x_{\beta_m}). \tag{25}$$

To simplify the notation, let each of the quantities in (24) equal $f'_\alpha(x_\alpha)$, and each of the quantities in (25) equal $f'_\beta(x_\beta)$.

Substituting (24) and (25) (which represent the induction hypothesis) into (22) and (23), the latter become:

$$f'_{\vec{\alpha}}(x_{\vec{\alpha}}) = f'_\alpha(x_\alpha) \sum_{\alpha_i \in \vec{\alpha}} \frac{x_{\alpha_i}}{\sum_{\alpha_j \in \vec{\alpha}} x_{\alpha_j}} \text{ and} \tag{26}$$

$$f'_{\vec{\beta}}(x_{\vec{\beta}}) = f'_\beta(x_\beta) \sum_{\beta_i \in \vec{\beta}} \frac{x_{\beta_i}}{\sum_{\beta_j \in \vec{\beta}} x_{\beta_j}}. \tag{27}$$

The summations on the RHSs of both of (26) and (27) can be trivially seen to sum to unity since they represent probabilities (in the conditioned spaces), implying that:

$$\forall i : f'_{\vec{\alpha}}(x_{\vec{\alpha}}) = f'_{\alpha_i}(x_{\alpha_i}) \text{ and} \tag{28}$$

$$\forall i : f'_{\vec{\beta}}(x_{\vec{\beta}}) = f'_{\beta_i}(x_{\beta_i}). \tag{29}$$

Combining the above with (21) yields:

$$f'_{\alpha_1}(x_{\alpha_1}) = \cdots = f'_{\alpha_m}(x_{\alpha_m}) = f'_{\beta_1}(x_{\beta_1}) = \cdots = f'_{\beta_m}(x_{\beta_m}), \tag{30}$$

implying that the global optimum required by the Lagrangian has been found. Hence, the theorem! □

Remarks. Theorem 2 has some very interesting consequences listed below:

1. The proof of Theorem 2 has tacitly assumed that all the automata have converged before the global convergence can be asserted. This implies that the TRAA $T_{d,j}$ is aware of its capacity, and that this is a known quantity to the TRAAs $T_{d+1,2j-1}$ and $T_{d+1,2j}$. In other words, if all the individual TRAAs converge to their local optimum, Theorem 2 states that the global optimum is attained. Conceptually, this can pose a small implementation-related problem. The fact is that the TRAAs of the lower level are converging even while the TRAA at the higher level is attempting to find *its* capacity. Therefore, essentially, the lower level TRAAs are working in a nonstationary environment. The strategy by which we can resolve this is to ensure that the higher level automata converge at a slower rate than the lower ones (thus, guaranteeing a certain level of stationarity). In practice, however, we have observed that if the resolution parameter N is large enough (in the order of hundreds) the time varying phenomenon is marginal, and the TRAAs at all the levels tend to converge simultaneously.

2. Theorem 2 claims that the solution obtained by the convergence of the individual TRAAs leads to the global convergence of the overall optimization problem. But this claim means that the ordering of the materials at the leaf nodes does not carry any significance. This is, indeed, true! It turns out that if the nodes at the leaves are ordered in such a way that "more precious materials" lie in the same subtree, the weight associated with the subtree of the composite material containing these "more precious materials" will have a much larger weight, and the weight of the other subtrees will be much smaller. As opposed to this, if the "more precious materials" lie in distinct subtrees, the weights associated with the respective subtrees will be correspondingly compensated for.

3 Empirical Results I: Linear and Exponential Unit-Value Functions

The H-TRAA solution has been rigorously tested for numerous cases and the solutions obtained have been, in our opinion, categorically remarkable. Its performance in terms of speed, convergence accuracy, and scalability have consistently been far more than we initially anticipated.

3.1 Problem Specification

In order to achieve a comprehensive test suite, we have conducted our experiments for two parametric objective functions (referred to as $E_i(x_i)$ and $L_i(x_i)$) being optimized.

These functions can be seen as representative for the class of concave objective functions that we address. We have also conducted experiments with a number of other objective functions, including those found in [10] and [11]. However, it turns out that $E_i(x_i)$ and $L_i(x_i)$ are particularly useful in the sense that they permit us to appropriately model a large range of distinct material unit-value functions. This, in turn, allows us to construct arbitrarily difficult and large knapsack problems, simply by means of a material index i. Furthermore, the findings presented here encompasses the findings obtained using other objective functions. Additionally, as will become clear below, the function $L_i(x_i)$ does not have a double derivative, and consequently, the feedback given will not be smoothly changing with x_i, as is typical in the applications we have considered [10], [11]. Thus, we believe that $E_i(x_i)$ and $L_i(x_i)$ are together sufficiently powerful to enable us to explore the performance of the H-TRAA for a diverse range of indexes, i.

More specifically, these objective functions have been given below for a material with index i as:

$$E_i = \frac{0.7}{i}(1 - e^{-ix_i}), \tag{31}$$

$$L_i = 0.7 \cdot x_i - \frac{1}{2} i \cdot x_i^2, \quad \text{If } x_i \leq \frac{0.7}{i}, \tag{32}$$

$$= \frac{0.7^2}{i}, \quad \text{If } x_i > \frac{0.7}{i}. \tag{33}$$

In the above, the constants are based on the boundary conditions due the contributions of x_i at the boundary values. These constants, however, are not crucial in the optimization because the corresponding unit-value functions are obtained as their respective derivatives. These are two probability functions given below for a material with index i as $E_i'(x_i)$ and $L_i'(x_i)$, respectively, which fall exponentially and linearly as per (34) and (35):

$$E_i'(x_i) = 0.7 \cdot e^{-i \cdot x_i}, \tag{34}$$

$$L_i' = \text{Max}[0.7 - i \cdot x_i, 0]. \tag{35}$$

To clarify how these functions work, consider the functions $E_i'(x_i)$. Then the relative profitability of material i decreases with x_i, its presence in the mixture, exponentially. Thus, if $x_2 = 0.3$ (i.e., material 2 fills 30 percent of the knapsack), the marginal profitability of increasing the amount of x_2 is $e^{-2 \cdot (0.3)} = e^{-0.6}$. Observe that with the notation, the profitability of materials that have a smaller index decreases *slower* than the profitability of materials that have a higher index. One should also observe that unlike the exponential function, the linear function, $L_i'(x_i)$ has an interesting peculiarity—the function for material i touches the X-axis at a finite point. This implies that the function being optimized is quadratic and that it attains a maximum at this point—after which it remains constant. Thus, after attaining this point, it is futile to add any additional quantity of material i.

Given the above considerations, our aim is to find $\vec{x}$, the amounts of the materials that have to be included in the knapsack so as to maximize its value. In the first instance, we aim to:

$$\begin{aligned} \text{maximize} \quad & E(\vec{x}) = \sum_1^n E_i(x_i), \text{ where } E_i(x_i) = \int_0^{x_i} p_i(u)du, \\ & \text{and } p_i(x_i) = E_i'(x_i), \end{aligned} \tag{36}$$

$$\text{subject to} \quad \sum_1^n x_i = c \quad \text{and} \quad \forall i \in \{1, \ldots, n\}, x_i \geq 0, \tag{37}$$

where $E_i'(x_i)$ is given by (34).

Similarly, in the second case we aim to:

$$\begin{aligned} \text{maximize} \quad & L(\vec{x}) = \sum_1^n L_i(x_i), \text{ where } L_i(x_i) = \int_0^{x_i} p_i(u)du, \\ & \text{and} \quad p_i(x_i) = L_i'(x_i), \end{aligned} \tag{38}$$

$$\text{subject to} \quad \sum_1^n x_i = c \text{ and } \forall i \in \{1, \ldots, n\}, \quad x_i \geq 0, \tag{39}$$

where $L_i'(x_i)$ is given by (35). Note that in general application domains, we may not be able to observe $f_i'(x_i)$ directly—examining a potential solution may be the only way to reveal the success of the chosen allocation.

3.2 H-TRAA Solution

In order to find an H-TRAA Solution to the above problem, we must define the Stochastic Environment that the LA are to interact with. As seen in Section 2, the Stochastic Environment consists of the unit volume value functions $\mathcal{F}' = \{f_1'(x_1), f_2'(x_2), \ldots, f_n'(x_n)\}$, which are unknown to

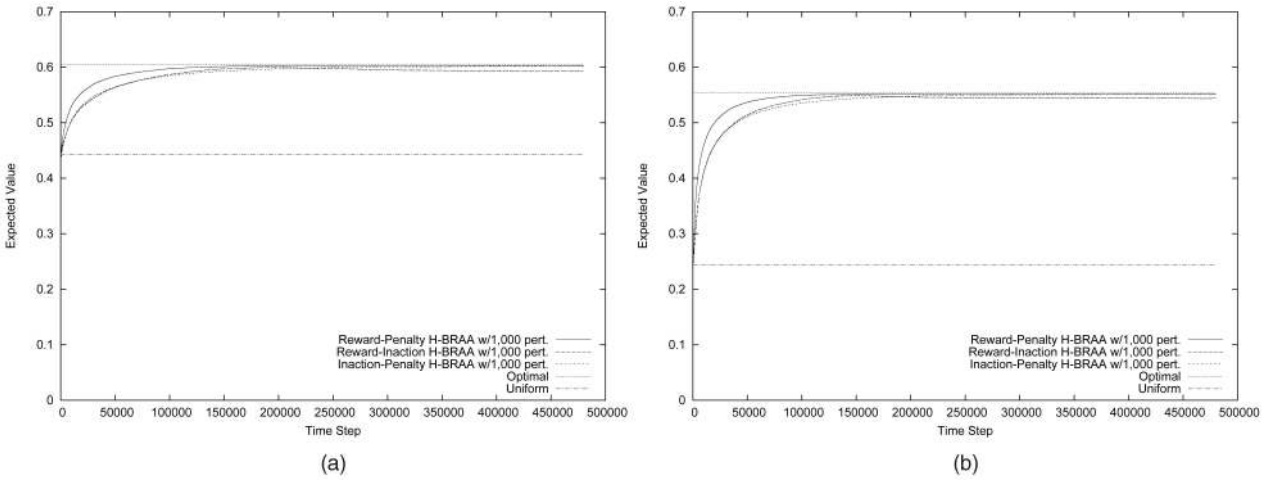

Fig. 5. Plots of the convergence of the H-TRAAs for the Exponential and Linear functions given by (34) and (35), respectively, for the Linear *Reward-Penalty*, *Inaction-Penalty*, and *Reward-Inaction* schemes. (a) The graph for the Exponential function, and (b) the graph for the Linear function. Using the same resolutions, we observe that the *Reward-Penalty* and the *Inaction-Penalty* updating schemes are the most accurate, even though a *Reward-Penalty* updating philosophy converges more quickly.

H-TRAA. We identify the nature of these functions by applying the principle of Lagrange multipliers to the above maximization problems. In short, after some simplification, it can be seen that the following conditions characterize the optimal solution for the exponential function of (34):

$$E_1'(x_1) = E_2'(x_2) = \cdots = E_n'(x_n),$$
$$\sum_1^n x_i = c \quad \text{and} \quad \forall i \in \{1, \ldots, n\}, \quad x_i \geq 0.$$

Similarly, after some simplification, it can also be seen that the following conditions characterize the optimal solution for the linear function of (35):

$$L_1'(x_1) = L_2'(x_2) = \cdots = L_n'(x_n),$$
$$\sum_1^n x_i = c \quad \text{and} \quad \forall i \in \{1, \ldots, n\}, \quad x_i \geq 0.$$

3.3 Experimental Setup

In this section, we evaluate our learning scheme by comparing it with three classical policies using synthetic data. We have implemented the following classical policies:

Uniform. The uniform policy allocates resources uniformly. This is the only classical policy that can be applied directly in an unknown environment.

Optimal. The optimal policy finds the optimal solution based on the principle of Lagrange multipliers [19], [27].

LAKG. The state-of-the-art solution for the nonlinear knapsack problem which is based on the so-called LAKG described in [9].

As we will see in the following, it turns that one of the strengths of the H-TRAA is its ability to take advantage of so-called spatial dependencies among materials. As mentioned earlier, in the above experimental setup, materials are spatially related in the sense that the updating probabilities decreases with the rank-index k. In order to starve the H-TRAA from this information, we opted to perturb this spatial structure. Each perturbation swapped the updating probabilities of a randomly selected material and the material succeeding it in the ranking. Based on the above, we conducted our experiments with 10^3, 10^4, 10^5, and 10^6 perturbations.

The results of our experiments are truly conclusive and confirm the power of the H-TRAA. Although numerous experiments were conducted (for various settings) and the number of automata, we report, for the sake of brevity, the results when the numbers of primitive materials was 512, 2,048, 8,192, and 32,768, and for the following environments:

- In the first case, the material unit-value function was the exponential function given by (34).
- In the second case, the material unit-value function was the linear function given by (35).

For these values, an ensemble of several independent replications with different random number streams was performed to minimize the variance of the reported results.

3.4 Configuring H-TRAA

The H-TRAA can be configured by various means. First of all, the material amount space $(0, 1)$ need not be discretized uniformly. Instead, a nonlinear material amount space can be formed, as done for the LAKG in [9]. Furthermore, the discretization resolution, N, must also be set for each TRAA, possibly varying from TRAA to TRAA in the hierarchy. In short, the performance achieved for a particular problem can be optimized using these different means of configuring the H-TRAA. In this section, however, our goal is to evaluate the overall performance of the H-TRAA, without resorting to fine tuning. Therefore, we will only use a linear material amount space, as specified in Section 2. Furthermore, in the experiments reported here, we have chosen to use the same resolution, $N = 2{,}000$, for all the TRAAs in the hierarchy, independent of the specific knapsack problem at hand. Thus, our aim is to ensure a fair comparison with the present state of the art, namely the LAKG scheme.

While the focus of the previous section was on learning only from material units of value 1 (rewards), with some simple modifications the H-TRAA scheme clearly supports the three well-established updating approaches:

1. **Reward-Inaction**: In this case, the H-TRAA updates its state only when a material unit volume value of "1" is given as the feedback from the Stochastic Environment, which is the case studied in the previous section.

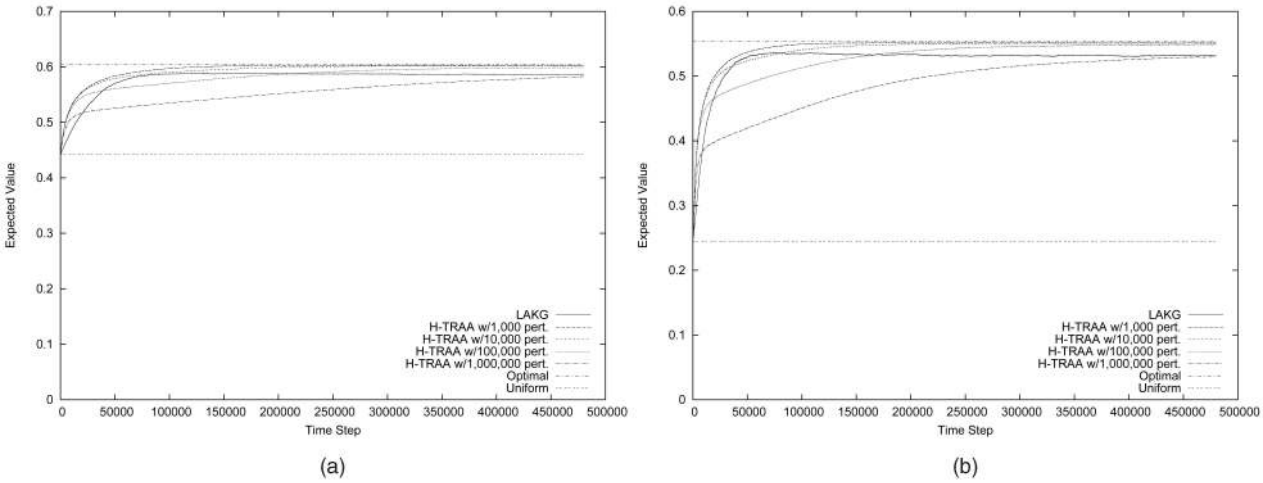

Fig. 6. Plots of the convergence of the H-TRAAs for the Exponential and Linear functions given by (34) and (35), respectively, for the Linear *Reward-Penalty* scheme. (a) The graph for the Exponential function, and (b) the graph for the Linear function. Under identical settings, we observe that the H-TRAA scheme is the most superior of the schemes.

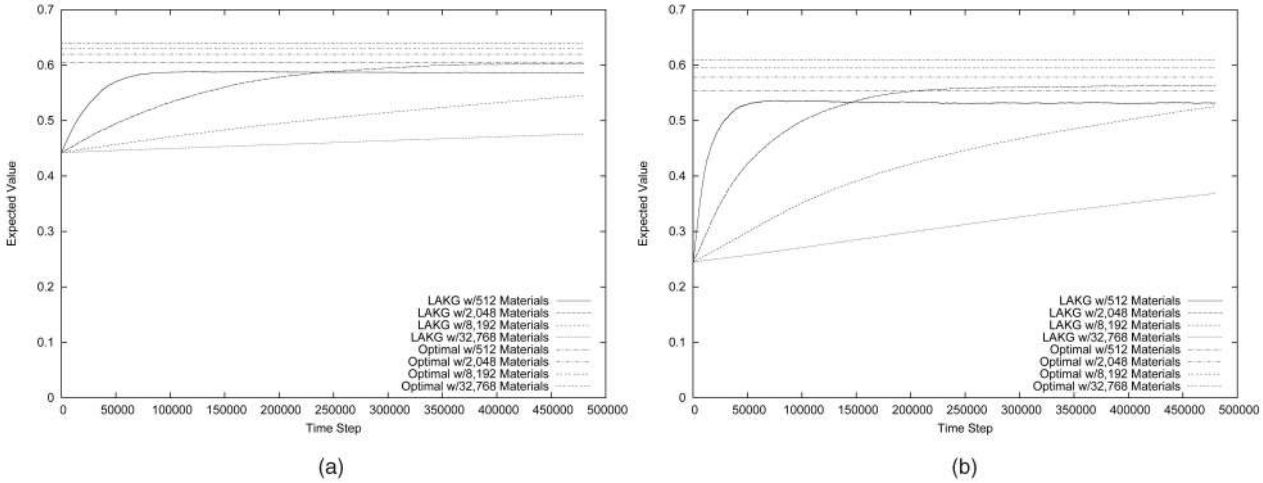

Fig. 7. Plots of the convergence of the LAKG for the Exponential and Linear functions given by (34) and (35) as the number of materials is increased significantly. (a) The graph for the Exponential function, and (b) the graph for the Linear function. Observe that the LAKG becomes prohibitively sluggish as the number of materials becomes very large.

2. **Inaction-Penalty**: In this case, the H-TRAA updates its state only when a material unit volume value of "0" is given as feedback. Here, the reader will observe that the state transitions of the individual TRAAs from Section 2 are inverted.
3. **Reward-Penalty**: In this case, the H-TRAA updates its state in both of the above cases.

From Fig. 5, we see that while interacting with the respective environments, the *Reward-Penalty* and *Inaction-Penalty* updating schemes are almost equiaccurate. However, the *Reward-Penalty* updating converges more quickly, since the state is updated both on rewards and on penalties. Further, merely relying on rewards to achieve updates seems to be slightly inferior to the other two approaches. Since we are emphasizing, in this paper, the speed of learning, we will, for the rest of this paper, utilize only the *Reward-Penalty* updating philosophies, even though the two other approaches produce almost similar results.

3.5 Static Environments

In this section, we present the results of the H-TRAA, LAKG, and the other schemes for static environments.

We see from Fig. 6 that the optimal policy provides a solution superior to the uniform policy solution. The figure also shows that the performance of the H-TRAA increases significantly quicker than the LAKG, the uniform and the optimal schemes. However, when the number of perturbations is increased, the performance of the H-TRAA tends to fall. Note that even with 10^6 perturbations, the H-TRAA provides a performance equal to the LAKG if each TRAA in the hierarchy is given a resolution N that is twice as large as the resolution applied by any of its children. Furthermore, the H-TRAA is more flexible than the LAKG, performing either better or similar to the H-TRAA configuration when it is optimized for the problem at hand. Observe too that the performance of both the H-TRAA and LAKG improve online (in a real-time manner) *without* invoking any parameter estimation phase. Both the LAKG and H-TRAA approach the performance boundary set by the optimal policy, and converge toward near-optimal solutions. However, the H-TRAA converges faster than the LAKG.

3.6 Scalability

One of the motivations for designing the H-TRAA was the improved scalability obtained by means of hierarchical learning. As seen in Fig. 7, extending the number of materials

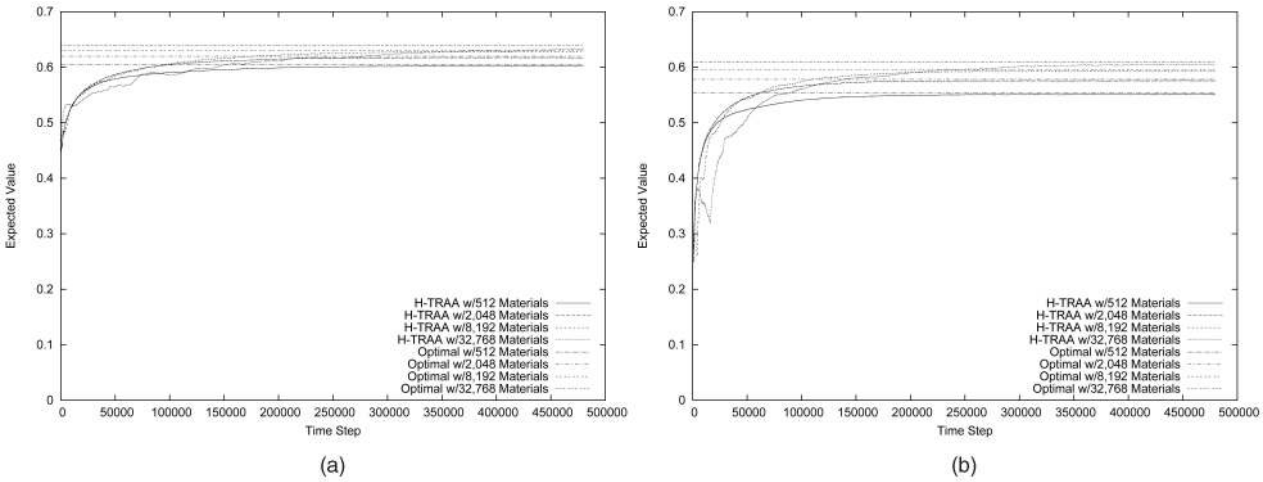

Fig. 8. Plots of the convergence of the H-TRAA for the Exponential and Linear functions given by (34) and (35) for the Linear *Reward-Penalty* scheme, as the number of materials is increased significantly. (a) The graph for the Exponential function, and (b) the graph for the Linear function. Observe that the H-TRAA scales sublinearly with the number of materials.

significantly increases the convergence time of LAKG. An increased initial learning phase may be unproblematic in cases where the system will run correspondingly longer, adapting to less dramatic changes as they occur. Fig. 8 displays the performance of the H-TRAA under identical settings. The reader should observe that increasing the number of materials does *not* significantly increase its convergence time. It is not at unfair to assert that under identical settings, the H-TRAA scheme is far superior to the LAKG. Indeed, the H-TRAA scales sublinearly with the number of materials.

4 Empirical Results II: Optimal Sample-Size Determination

To demonstrate the potential of the solution for practical applications, in this section, we consider the problem of estimating the proportion of a population possessing some specific characteristics. Specifically, we assume that n populations are to be evaluated, and that each population i is characterized by an independent unknown binomial proportion u_i. We will here pursue the goal of minimizing the variance of the proportion estimates when the total number of samples available for estimating the proportions is restricted to c. The purpose is to make the estimates as accurate as possible. As mentioned earlier, for instance, the task at hand could be to determine the proportion of a website that is successfully validated by an HTML validator [23], and that n websites are to be evaluated by only accessing c webpages.

4.1 Problem Specification

Let x_i be the number of elements sampled randomly from population i and let the count Y_i be the number of the sampled elements that possess a chosen characteristic. For large x_i and when u_i is not too near 0 or 1, the estimator $\hat{u}_i = \frac{Y_i}{x_i}$ is approximately normal with mean u_i and standard deviation [1]

$$s_i = \sqrt{\frac{u_i(1-u_i)}{x_i}}.$$

As seen, the standard deviation can be reduced (and the estimate accuracy increased) by increasing the number of samples x_i. In the problem targeted in this section, n different populations can be sampled c times and the goal is to distribute the samples among the populations to minimize the aggregated variance of the estimates. The problem can be reformulated as follows:

$$\begin{aligned} &\text{maximize} && \sum_{i=1}^{n} -\frac{u_i(1-u_i)}{x_i}, \\ &\text{subject to} && \sum x_i = c, \\ &1 \le x_i,\ i = 1, \ldots, n. \end{aligned}$$

The above optimization problem is an NEFK problem with concave and separable objective function. Since the u_is are assumed unknown, we apply our H-TRAA to find a near-optimal solution incrementally and online.

4.2 The H-TRAA Solution

We must first define the Stochastic Environment that the H-TRAA is to interact with. That is, we must define the stochastic functions $\mathcal{F}' = \{f'_1(x_1), f'_2(x_2), \ldots, f'_n(x_n)\}$. By applying the principles of Lagrange multipliers, we find the following conditions that characterize the optimal solution:

$$\begin{aligned} &\frac{u_1(1-u_1)}{{x_1}^2} = \cdots = \frac{u_n(1-u_n)}{{x_n}^2}, \\ &\sum x_i = c, \\ &1 \le x_i,\ i = 1, \ldots, n. \end{aligned}$$

Accordingly, we define $f'_i(x_i)$ as follows: First of all, each time $f'_i(x_i)$ is accessed by H-TRAA, population i is sampled once and the proportion estimate $\hat{u}_i$ is updated accordingly.[7] After $\hat{u}_i$ has been updated, we instantiate $f'_i(x_i)$ by a random draw—$f'_i(x_i)$ is instantiated to the value 0 with probability

$$1 - \frac{\hat{u}_i(1-\hat{u}_i)}{{x_i}^2}$$

7. For a dynamic environment, we would utilize a "window-based" strategy and only use the last c samples to estimate the u_is. However, we are currently studying how recently proposed weak estimators can be used in this setting [18].

TABLE 1
The True Population Proportions Used in the Experiment, and the Number of Populations Associated with Each Proportion

True Proportion	Populations
0.5	6
0.750 / 0.250	5
0.900 / 0.100	41
0.990 / 0.010	51
0.999 / 0.001	409

and to the value 1 with probability $\frac{\hat{u}_i(1-\hat{u}_i)}{x_i^2}$. In other words, we keep running estimates of the u_is in order to calculate the outcome probabilities of the $f_i'(x_i)$s.[8]

4.3 Experimental Setup

In this section, we evaluate our learning scheme by comparing it with the optimal and uniform policies using synthetic data. The reader should appreciate that, in practice, we can only apply the uniform policy, because the optimal policy requires that the u_is are known.

The data used in the experiment are summarized in Table 1. The table shows the true population proportions used and the number of populations associated with each proportion. The experiment encompasses 512 populations, and the corresponding proportions are to be estimated by allocating 50,000 samples.

4.4 Static Environments

We first report the results obtained for environments which are static, i.e., where the u_is, although unknown, do not change with time. As we will see in the following, it turns that one of the strengths of the H-TRAA is its ability to take advantage of so-called spatial dependencies among materials. As mentioned earlier, in the above experimental setup, materials are spatially related in the sense that the updating probabilities decreases with the rank-index k. In order to starve the H-TRAA from this information, we opted to perturb this spatial structure. Each perturbation swapped the probabilities of a randomly selected material and the material succeeding it in the ranking.

We conducted our experiments with 10^3, 10^4, 10^5, and 10^6 perturbations. For each of these values, an ensemble of several independent replications with different random number streams was performed so as to minimize the variance of the reported results. The results of our experiments are truly conclusive and confirm the power of the H-TRAA. Although several experiments were conducted using various setting for various numbers of automata, we report, in the interest of brevity, a brief overview of the results obtained.

Fig. 9 plots the variance of the current solution (as a function of time) each time a unit volume value function $f_i'(x_i)$ has been sampled. The graphs show the results of applying the H-TRAA with 5,000 states and the LAKG with 12,500 states (where the amount of the material added on a transition in the latter is not fixed but varying in a nonlinear manner[9]).

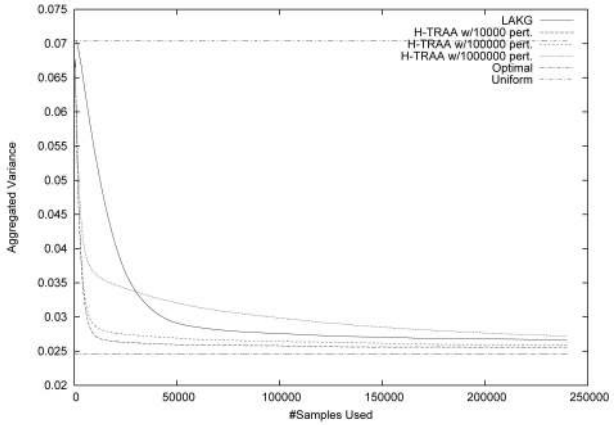

Fig. 9. The performance of LAKG and H-TRAA in the static environment for the optimal sample-size determination problem.

As seen in the figure, the H-TRAA steadily reduces the variance of the initial solution in which the populations are sampled uniformly. Indeed, even by the first 50,000 samples, one can observe a very *significant* reduction. The reader should notice that the H-TRAA converges to a near-optimal allocation more expediently and far quicker than the LAKG scheme, expect for the case with 1,000,000 perturbations where the H-TRAA initially converges faster but subsequently in a more conservative manner.

Also note that the confidence interval of each estimated proportion is reduced by minimizing the total variance.

4.5 Scalability

Again, as seen in Fig. 10, extending the number of materials increases the convergence time of the LAKG. The H-TRAA, however, is far less affected by the number of materials.

From Fig. 10, we observe that while the LAKG does not even converge (see the figure on the left), the H-TRAA scales sublinearly *in every case* with the number of materials (the figure on the right). However, the most interesting phenomenon that we observe from Fig. 10 is the ability of the H-TRAA to emerge out of local optima. The H-TRAA first decreases to a minimum, but when it "discovers" that there is a better solution (which in this case implies a superior partitioning of the nodes in the tree to their left and right subtrees), it is capable of unlearning the inferior configuration and converging to a superior solution. This, we believe, is quite remarkable, especially because the size of the underlying tree is very large, implying that the number of possible binary trees (which grows exponentially with the size) is even larger. However, by arriving at the global optimum, we see that the H-TRAA has succeeded in learning the best tree structure to resolve the sampling proportions!

5 Conclusions and Further Work

In this paper, we have considered the fractional knapsack problem and extended the non-LA state of the art in two

8. Because the outcome probabilities are always available for the populations, we can normalize the outcome probabilities to speed up convergence.

9. The details of this are omitted. They can be found in [9].

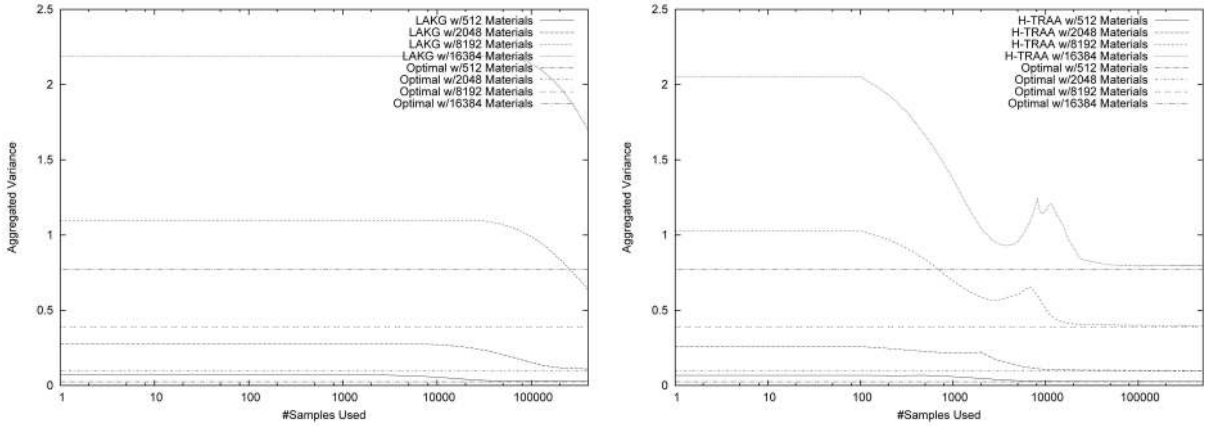

Fig. 10. Scalability of LAKG and H-TRAA for the optimal sample-size determination problem.

ways. First of all, we have treated the unit volume values of each material as a *stochastic* variable whose distribution is *unknown*. Second, we have worked with the model that the expected value of a material may decrease after each addition to the knapsack. The learning scheme we proposed for solving this knapsack problem was based on a hierarchy of so-called TRAA. Each TRAA works with two materials and moves along the probability space discretized by a resolution parameter, N, with a random walk whose coefficients are not constant. The asymptotic optimality of the TRAA has been proven. We have then presented a formal theory by which an ensemble of TRAAs (the H-TRAA) can be structured in a hierarchical manner to deal with a very large number of materials, and its asymptotic optimality has also been proven. Comprehensive experimental results have demonstrated that performance of the H-TRAA is superior to the previous state-of-the-art schemes, including the so-called, LAKG [9]. Additionally, for a given precision, our scheme determines the material fractions of maximal expected value by invoking online interactions with the knapsack. We have also provided empirical evidence to clearly demonstrate the H-TRAAs sublinear scaling property.

In our further work, we first of all intend to utilize the H-TRAA solution to resolve the web-polling problem, and the problem of determining the optimal sample size required for estimation purposes. Some preliminary results for this are already available, and will be presented elsewhere [10], [11]. We also aim to develop alternate LA-based solutions for different classes of knapsack problems, including the NP-hard integer knapsack problem. Essentially, we propose to do this by enhancing the concepts introduced in this paper with a branch-and-bound-based relaxation capability. Finally, we are also currently investigating how other classes of LA can form the basis for solving knapsack-type problems.

ACKNOWLEDGMENTS

The work of B. John Oommen was partially supported by the Natural Sciences and Engineering Research Council of Canada. He is also an adjunct professor with the University of Agder in Grimstad, Norway.

REFERENCES

[1] G.K. Bhattacharyya and R.A. Johnson, *Statistical Concepts and Methods.* John Wiley and Sons, 1977.

[2] P.E. Black, "Fractional Knapsack Problem," *Dictionary of Algorithms and Data Structures,* 2004.

[3] K.M. Bretthauer and B. Shetty, "The Nonlinear Knapsack Problem—Algorithms and Applications," *European J. Operational Research,* vol. 138, pp. 459-472, 2002.

[4] B.C. Dean, M.X. Goemans, and J. Vondrak, "Approximating the Stochastic Knapsack Problem: The Benefit of Adaptivity," *Proc. 45th Ann. IEEE Symp. Foundations of Computer Science (FOCS '04),* pp. 208-217, 2004.

[5] B. Fox, "Discrete Optimization via Marginal Analysis," *Management Sciences,* vol. 13, no. 3, pp. 211-216, 1966.

[6] M.R. Gibson, J.W. Ohlmann, and M.J. Fry, "An Agent-Based Stochastic Ruler Approach for a Stochastic Knapsack Problem with Sequential Competition," *Computers and Operations Research,* vol. 37, no. 3, pp. 598-609, 2010.

[7] O.-C. Granmo and B.J. Oommen, "On Allocating Limited Sampling Resources Using a Learning Automata-Based Solution to the Fractional Knapsack Problem," *Proc. 2006 Int'l Intelligent Information Processing and Web Mining Conf. (IIS:IIPW '06),* pp. 263-272, 2006.

[8] O.-C. Granmo, B.J. Oommen, S.A. Myrer, and M.G. Olsen, "Determining Optimal Polling Frequency Using a Learning Automata-Based Solution to the Fractional Knapsack Problem," *Proc. 2006 IEEE Int'l Conf. Cybernetics and Intelligent Systems (CIS) and Robotics, Automation and Mechatronics (RAM),* pp. 73-79, 2006.

[9] O.-C. Granmo, B.J. Oommen, S.A. Myrer, and M.G. Olsen, "Learning Automata-based Solutions to the Nonlinear Fractional Knapsack Problem with Applications to Optimal Resource Allocation," *IEEE Trans. Systems, Man, and Cybernetics, Part B,* vol. 37, no. 1, pp. 166-175, Feb. 2007.

[10] O.-C. Granmo and B.J. Oommen, "A Hierarchy of Twofold Resource Allocation Automata Supporting Optimal Web Polling," *Proc. 21st Int'l Conf. Industrial and Eng. Applications of Artificial Intelligence and Expert Systems (IEA/AIE '08),* pp. 347-358, 2008.

[11] O.-C. Granmo and B.J. Oommen, "A Hierarchy of Twofold Resource Allocation Automata Supporting Optimal Sampling," *Proc. 22nd Int'l Conf. Industrial and Eng. Applications of Artificial Intelligence and Expert Systems (IEA/AIE '09),* pp. 523-534, 2009.

[12] J. Kangasharju, F. Hartanto, M. Reisslein, and K.W. Ross, "Distributing Layered Encoded Video through Caches," *IEEE Trans. Computers,* vol. 51, no. 66, pp. 622-636, June 2002.

[13] H. Kellerer, U. Pferschy, and D. Pisinger, *Knapsack Problems.* Springer, 2004.

[14] G.Y. Lin, Y. Lu, and D.D. Yao, "The Stochastic Knapsack Revisited: Structure, Switch-Over Policies, and Dynamic Pricing," *Operations Research,* vol. 56, pp. 945-957, 2008.

[15] K.S. Narendra and M.A.L. Thathachar, *Learning Automata: An Introduction.* Prentice Hall, 1989.

[16] B.J. Oommen, "Absorbing and Ergodic Discretized Two Action Learning Automata," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 16, no. 2, pp. 282-293, Mar./Apr. 1986.

[17] B.J. Oommen, "Stochastic Searching on the Line and Its Applications to Parameter Learning in Nonlinear Optimization," *IEEE Trans. Systems, Man, and Cybernetics, Part B,* vol. 27, no. 4, pp. 733-739, Aug. 1997.
[18] B.J. Oommen and L. Rueda, "Stochastic Learning-Based Weak Estimation of Multinomial Random Variables and Its Applications to Pattern Recognition in Non-Stationary Environments," *Pattern Recognition,* vol. 39, pp. 328-341, 2006.
[19] S. Pandey, K. Ramamritham, and S. Chakrabarti, "Monitoring the Dynamic Web to Respond to Continuous Queries," *Proc. 12th Int'l World Wide Web Conf.,* pp. 659-668, 2003.
[20] T.C. Perry and J.C. Hartman, "An Approximate Dynamic Programming Approach to Solving a Dynamic, Stochastic Multiple Knapsack Problem," *Int'l Trans. Operational Research,* vol. 16, no. 3, pp. 347-359, 2009.
[21] K.W. Ross and D.H.K. Tsang, "The Stochastic Knapsack Problem," *IEEE Trans. Comm.,* vol. 37, no. 7, pp. 740-747, July 1989.
[22] J. Sachs, "A Stochastic Knapsack Model for the Capacity Evaluation of (Multi-) Radio Access Networks," *Int'l J. Systems Science,* vol. 32, no. 3, pp. 79-87, 2006.
[23] M. Snaprud, N. Ulltveit-Moe, O.-C. Granmo, M. Rafoshei-Klev, A. Wiklund, and A. Sawicka, "Quantitative Assessment of Public Web Sites Accessibility—Some Early Results," *Proc. Accessibility for All Conf.,* 2003.
[24] E. Steinberg and M.S. Parks, "A Preference Order Dynamic Program for a Knapsack Problem with Stochastic Rewards," *J. Operational Research Soc.,* vol. 30, no. 2, pp. 141-147, 1979.
[25] M.A.L. Thathachar and P.S. Sastry, *Networks of Learning Automata: Techniques for Online Stochastic Optimization.* Kluwer Academic Publishers, 2004.
[26] M.L. Tsetlin, *Automaton Theory and Modeling of Biological Systems.* Academic Press, 1973.
[27] J.L. Wolf, M.S. Squillante, J. Sethuraman, and K. Ozsen, "Optimal Crawling Strategies for Web Search Engines," *Proc. 11th Int'l World Wide Web Conf.,* pp. 136-147, 2002.

Ole-Christoffer Granmo received the MSc degree in 1999 and the PhD degree in 2004, both from the University of Oslo, Norway. He is currently an associate professor in the Department of ICT, University of Agder, Norway. His research interests include intelligent systems, stochastic modeling and inference, machine learning, pattern recognition, learning automata, distributed computing, and surveillance and monitoring. He is the author of more than 34 refereed journal and conference publications.

B. John Oommen received the BTech degree from the Indian Institute of Technology, Madras, in 1975. He obtained the ME degree from the Indian Institute of Science, Bangalore, in 1977. He then went on for the MS and PhD degrees which he received from Purdue University, in West Lafayettte, Indiana, in 1979 and 1982, respectively. He joined the School of Computer Science at Carleton University in Ottawa, Canada, in the 1981-1982 academic year. He is still at Carleton and holds the rank of a full professor. Since July 2006, he has been awarded the honorary rank of Chancellor's Professor, which is a lifetime award from Carleton University. His research interests include automata learning, adaptive data structures, statistical and syntactic pattern recognition, stochastic algorithms, and partitioning algorithms. He is the author of more than 320 refereed journal and conference publications, and is a fellow of the IEEE and the IAPR. He has been on the editorial board of the *IEEE Transactions on Systems, Man, and Cybernetics* and the *Pattern Recognition*.