Optimal Node Visitation in Acyclic Stochastic Digraphs with Multi-threaded Traversals and Internal Visitation Requirements*

Theologos Bountourelis and Spyros Reveliotis
School of Industrial & Systems Engineering
Georgia Institute of Technology
{tbountou,spyros}@isye.gatech.edu

Abstract

The original definition of the problem of optimal node visitation (ONV) in acyclic stochastic digraphs concerns the identification of a routing policy that will enable the visitation of each *leaf* node a requested number of times, while minimizing the expected number of the graph traversals. The original work of [1] formulated this problem as a Stochastic Shortest Path (SSP) problem, and since the state space of this SSP formulation is exponentially sized with respect to the number of the target nodes, it also proposed a suboptimal policy that is computationally tractable and asymptotically optimal. This paper extends the results of [1] to the cases where (i) the tokens traversing the graph can *"split"* during certain transitions to a number of (sub-)tokens, allowing, thus, the satisfaction of many visitation requirements during a single graph traversal, and (ii) there are additional visitation requirements attached to the *internal* graph nodes, which, however, can be served only when the visitation requirements of their successors have been fully met. In addition, the presented set of results establishes stronger convergence properties for the proposed suboptimal policies, and it provides a formal complexity analysis of the considered ONV formulations. From a practical standpoint, the extension of the original results performed in this paper enables their effective usage in the application domains that motivated the ONV problem, in the first place.

1 Introduction

The ONV problem and its practical motivation The problem of the optimal node visitation (ONV) in acyclic stochastic digraphs was originally introduced in [1]. According to the definition provided in [1], this problem concerns the identification of a routing policy that will enable the visitation of each leaf node of an acyclic stochastic digraph a requested number of times, while minimizing the expected number of the graph traversals. In [1], the problem was formulated as a *stochastic shortest path (SSP)* problem [2], and due to the state space explosion in the provided SSP formulation, it was eventually addressed by a suboptimal policy that traded off optimality for computational tractability. This policy

*An abridged version of this manuscript was presented at WODES'08.

was derived from a continuous – or *"fluid"* – relaxation of the original problem, and it was shown to be *asymptotically optimal*, in the sense that the ratio of its performance to the performance of an optimal policy converges to one, as the node visitation requirements are scaled uniformly to infinity.

From a more practical standpoint, the ONV problem mentioned above was motivated by our past work presented in [3, 4]. In these works, a learning agent must compute on-line an optimal policy for a task that evolves episodically over a state space that is acyclic, and it has a single source state that defines the task initial state. The execution of an action implies an immediate stochastic reward and a stochastic transition to a subsequent state. However, the statistics of the collected rewards and the various transition probabilities are initially unknown to the learning agent. Furthermore, the considered task can involve *multi-threading*, with the different threads being initiated upon the execution of certain actions at the visited states. The objective of the learning agent is to maximize the expected value accumulated over a single episode, through the selection of a pertinent action at each state visited by each running thread.

In [4], it is shown that the agent can obtain an ϵ-optimal policy with probability at least $1 - \delta$, by sampling the various actions available at each task state a certain number of times[1] and selecting the action that results to the highest sample mean. The algorithm's sampling schedule essentially constitutes a set of pre-specified *visitation requirements* for each task state. Furthermore, any viable schedule for preforming these visitations must observe the additional requirement that a task state can have the value of its local actions sampled only if all of its successors states have been fully sampled and the sought policy has been determined at these states. Hence, at any point of the algorithm execution, the various task states are naturally classified as "inactive", "actively explored" and "fully explored", and the considered algorithm can be summarized as follows: Starting with the set of terminal states, the algorithm maintains a "frontier" set of "actively explored" states for which it tries to identify an optimal action. When an "actively explored" state has been visited a number of times equal to the respective visitation requirement and, thus, all of its local actions have been adequately sampled, it is assigned the action with the highest sample mean, it is declared "fully explored", and it abandons the set of "actively explored" states. On the other hand, "inactive" states join the "frontier" layer of "actively explored" states when all their immediate successors become "fully explored".

At every task iteration – or task *episode* – executed by the algorithm described in the previous paragraph, the learning agent has to navigate all the activated threads through a contiguous set of unexplored states until they reach an actively explored or a fully explored state. When an activated thread reaches an actively explored state, the algorithm collects a sample regarding the value of one of the state actions, and it reduces the state visitation requirement by one unit. On the other hand, a thread that results in a fully explored state,[2] does not contribute any additional information to the algorithm's sampling process. It is clear from this description that, in order to effect an expedient learning process, there is a need for pertinent routing policies for the threads activated at each task episode, that will enable the realization of the posed visitation requirements in a minimum number of episodes. Furthermore, such

[1] that depends on the graph structure and the performance parameters ϵ and δ

[2] This can happen due to the stochastic nature of the task transitions.

an optimized routing policy must base the agent's decisions on (i) the current set of inactive, actively explored and fully explored states, (ii) the set of the visitation requirements remaining for each state, (iii) the states of all the activated threads, and (iv) the probability distributions that govern the stochastic outcomes of the different actions that can be exerted by the activated threads. The ONV problem of [1] and its new variations that are studied in the later parts of this work, constitute a series of *prototypical abstractions* of the aforementioned routing problem, of increasing modeling detail and corresponding complexity. In an effort to develop the necessary analytical insights and a theoretical framework for the effective design of the sought routing policies, all of the provided formulations assume a state of *"perfect information"* for the routing agent; in particular, all of these formulations incorporate the simplifying assumption that all the transition probabilities of the underlying task are known *a priori*. However, the policies derived from this analysis can be subsequently implemented in the context of the learning algorithm described above, according to a *"certainty equivalence"* scheme [2] that substitutes the actual transition probabilities with pertinent estimates obtained during the execution of the algorithm.

The paper contributions Next we detail the major contributions of this work with respect to the ONV problem that was motivated and outlined in the previous paragraph. As it will be revealed from the following discussion, the presented results enhance the modeling affinity and the relevance of the ONV problem to its motivational application context, and they also strengthen the theoretical analysis of the underlying problem dynamics in a way that facilitates the design of more pertinent solutions to it. A third line of the results presented in this manuscript concerns the systematic investigation of the computational complexity of the considered variations of the ONV problem, a task that provides formal testimony to their increased (non-polynomial) complexity, but also reveals the affinity of the considered problems to some more classical stochastic scheduling problems that have been addressed in the literature. A more detailed account of these contributions is as follows:

We start with a discussion of the way that the presented results increase the modeling affinity and the relevance of the ONV problem to the machine learning context that motivated it. As stated in the opening paragraph of this section, the ONV formulation addressed in [1] considered the case where the only nodes possessing non-zero visitation requirements are the terminal nodes of the underlying stochastic digraph. Furthermore, this first formulation did not consider any multi-threading effects in its analysis. In this work, building upon the insights and the results obtained in [1], we take on the additional features of task multi-threading and the presence of non-zero visitation requirements at the non-terminal nodes. In particular, we attempt this extension in two steps, with the first step introducing and analyzing the effects of multi-threading, and with the second step employing the results of the first in order to address the more complex problem version that results from the addition of internal visitation requirements. It is shown that, similar to the original ONV problem version of [1], both of these new variations of the ONV problem can be modeled as SSP problems that suffer from a state space explosion. Hence, for both of these new cases, fluid relaxations are proposed that can lead to randomized policies that are computationally tractable and present good performance with respect to the performance of the corresponding optimal policies. More specifically, the randomized policy developed for the case of the ONV problem with task multi-threading but without internal visitation requirements remains

asymptotically optimal under a uniform scaling to infinity of the posed visitation requirements. On the other hand, the optimization problem defined by the fluid relaxation of the ONV variation that contains, both, task multi-threading and internal visitation requirements, is a complex hybrid optimal control problem of limited computational tractability [5]. Hence, in order to obtain computationally efficient policies for this ONV variation, we confine our analysis within a class of randomized policies that are easily implementable, and we provide a fluid relaxation that leads to a policy which is asymptotically optimal within the scope of the considered policies.

From a more technical standpoint, the presented work expands the methodology developed in [1] for the development of efficient suboptimal policies for the considered ONV problem, by basing the relevant analyses on renewal theory [6], instead of the strong law of large numbers that was used in [1]. As a result, the provided analyses are also able to establish bounds for the divergence of the performance of the aforementioned policies from the performance of the corresponding optimal policy, as the posed visitation requirements are scaled to infinity. Even more interestingly, this new line of analysis has also revealed a number of cases of considerable practical significance where the aforementioned divergence is uniformly bounded by a constant.

Finally, as mentioned above, another line of investigation of the ONV variations considered in this work concerns the formal study of the computational complexity of these problems. Along this line, it is established that (i) the introduction of the multi-threading effect in the ONV problem renders it PSPACE-hard [7], while (ii) the presence of internal visitation requirements makes it at least as hard as the *"Poisson-tree" scheduling* problem [7], a stochastic scheduling problem whose computational complexity is an open issue. The derivation of this last result reveals also the structural similarities of the considered ONV variations to some other stochastic scheduling problems previously studied in the literature.

Indeed, we should notice, at this point, that our analysis for the ONV problem outlined in the previous paragraphs is similar, in spirit, to the prevailing trends regarding the analysis of stochastic scheduling problems [8, 9]. As indicated in [8], most stochastic scheduling problems are notoriously hard to solve optimally, and one has to compromise for solutions that are suboptimal but computationally tractable. In particular, the last few years have seen the emergence of a number of works that seek to provide suboptimal solutions to various stochastic scheduling problems by exploiting some "relaxed" – or "fluid"-based – version of the original problem [10, 11, 12]. Furthermore, in many cases, this line of analysis also provides guaranteed bounds for the potential suboptimality of the derived policies; cf., for instance, the works of [13, 14] and the references provided therein. However, it is also true that the application of such a research program to any given stochastic scheduling problem is a major challenge in itself, since the detailed results, their supportive arguments and their implementational complexity are strongly dependent on the particular structure and attributes of the considered problem; the recent publication of [12] provides an excellent exposition of (most of) the relevant theory and testifies to these effects.

The rest of the paper is organized as follows: Section 2 introduces the variation of the ONV problem with multi-threaded traversals, establishes its PSPACE-hardness, and provides an asymptotically

optimal randomized policy for it. Section 3 introduces the problem variation which further allows for the assignment of visitation requirements to non-terminal nodes, and extends the results developed in Section 2 to this new problem case. In addition, Section 3 discusses the relationship of this new version of the ONV problem to the "Poisson-tree" scheduling problem mentioned above. Finally, Section 4 concludes the paper by summarizing its major developments, and highlighting directions for future work.

2 The ONV problem with multi-threaded traversals

A formal description of the considered ONV problem An instance of the problem considered in this section is completely defined by a quadruple $\mathcal{E} = (X, \mathcal{A}, \mathcal{P}, \mathcal{N})$, where

- X is a finite set of *nodes*, that is partitioned into a sequence of *"layers"*, X^0, X^1, ..., X^L. $X^0 = \{x^0\}$ defines the *source* or *root node*, while nodes $x \in X^L$ are the *terminal* or *leaf* nodes.

- $\mathcal{A}$ is a set function defined on X, that maps each $x \in X$ to the finite, non-empty set $\mathcal{A}(x)$, comprising all the *decisions / actions* that can be executed by the control agent at node x. It is further assumed that for $x \neq x'$, $\mathcal{A}(x) \cap \mathcal{A}(x') = \emptyset$.

- $\mathcal{P}$ is the *transition function*, defined on $\bigcup_{x \in X \setminus X^L} \mathcal{A}(x)$, that associates with every action a in this set a discrete probability distribution $p(\cdot; a)$. The support sets, $\mathcal{S}(a)$, of the distributions $p(\cdot; a)$ consist of *multi-sets* of X [3] that satisfy the following property: For any given action $a \in \mathcal{A}(x)$ with $x \in X^i$ for some $i = 0, \ldots, L-1$, the multi-sets in $\mathcal{S}(a)$ have their elements constrained in $\bigcup_{j=i+1}^{L} X^j$. In the motivational context of the ONV problem that was discussed in the introductory section, each multi-set $\nu_{a,k} \in \mathcal{S}(a)$, $k = 1, \ldots |\mathcal{S}(a)|$, implies the *"splitting"* of the thread that executes action a, into a number of sub-threads equal to the total number of elements in $\nu_{a,k}$. More specifically, for every $i = 1, \ldots |X|$, $\nu_{a,k}(i)$ of these sub-threads are initialized at the task state corresponding to component $\nu_{a,k}(i)$. On the other hand, the requirement that for any $a \in \mathcal{A}(x)$ and $x \in X^i$ the multi-sets of $\mathcal{S}(a)$ are constrained in $\bigcup_{j=i+1}^{L} X^j$, is a formal statement of the acyclic structure of the underlying task dynamics over a single episode. We also notice, at this point, that, for the subsequent developments, it is more intuitive to think of the various active threads that evolve in task state space X, as *"tokens"* that are traversing X. In this new interpretation, the execution of an action $a \in \mathcal{A}(x)$ by a token located in node x results in its substitution by one of the multi-sets of tokens in the support set $\mathcal{S}(a)$, according to the distribution $p(\cdot; a)$.

- $\mathcal{N}$ is the *visitation requirement vector*, that associates with each node $x \in X^L$ a visitation requirement $\mathcal{N}_x \in \mathbf{Z}_0^+$. The *support* $||\mathcal{N}||$ of $\mathcal{N}$ is defined by the nodes $x \in X^L$ with $\mathcal{N}_x > 0$; we shall refer to nodes $x \in ||\mathcal{N}||$ as the problem *"target"* nodes.

- Finally, we define the *instance size* $|\mathcal{E}| \equiv |X| + |\bigcup_{x \in X} \mathcal{A}(x)| + |\mathcal{N}|$, where application of the operator $|\,|$ on a set returns the cardinality of this set, while application on a vector returns its l_1 norm.

[3] We remind the reader that a multi-set defined on a set X is essentially a vector ν of dimensionality $|X|$ and with elements belonging to $\mathbf{Z}_0^+$, the set of non-negative integers. Each component $\nu(i)$ of vector ν corresponds to one of the elements of X and its value indicates how many replicates of this element are included in the multi-set represented by ν.

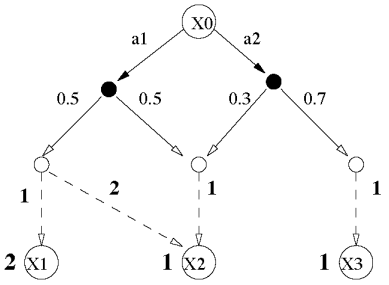

Figure 1: The stochastic graph for the problem instance considered in Example 1.

In the subsequent discussion we shall employ the variable vector $\mathcal{N}^c$ to denote the *vector of the remaining visitation requirements*. The control agent starts at period $t = 0$, by placing a *token* at node x^0 and setting $\mathcal{N}^c := \mathcal{N}$. At every consecutive period $t = 1, 2, 3, \ldots$, it (i) observes the current *configuration* g, i.e. the number and position of the tokens in the set $X \setminus X^L$ and the vector of remaining visitation requirements, $\mathcal{N}^c$, (ii) selects an action $a \in \mathcal{A}(x)$ and commands its execution on a single token at node x, (iii) generates the new tokens at the nodes indicated by the multi-set selected according to the probabilities $p(\cdot, a)$, (iv) updates $\mathcal{N}^c_x$ to $(\mathcal{N}^c_x - k)^+$ when k tokens reach one of the terminal nodes, $x \in X^L$, and finally, when the last token exits the set $X \setminus X^L$, (v) *resets* itself by placing a token at the initial node x^0, in order to start another traversal. The entire operation terminates when all the node visitation requirements have been reduced to zero. Our intention is to determine an action selection scheme – or a *policy* – π, that maps each configuration g to an action $\pi(g) \in \bigcup_{x \in X \setminus X^L} \mathcal{A}(x)$ in a way that minimizes the expected number of graph traversals until $\mathcal{N}^c = \mathbf{0}$. The set of all possible policies for the considered problem will be denoted by Π.

Example 1 As an example, we consider the problem instance depicted in Figure 1. In this case, there are two actions, a^1 and a^2, emanating from the root node x^0, and three leaf nodes, x^1, x^2 and x^3. Ordering the nodes of X in increasing order with respect to their ID number, the set $\mathcal{S}(a^1)$ consists of the two multi-sets $\nu_{a^1,1} = [0, 1, 2, 0]$ and $\nu_{a^1,2} = [0, 0, 1, 0]$, whereas the set $\mathcal{S}(a^2)$ consists of the multi-sets $\nu_{a^2,1} = [0, 0, 1, 0]$ and $\nu_{a^2,2} = [0, 0, 0, 1]$. Furthermore, $p(\nu_{a^1,1}; a^1) = 0.5$, $p(\nu_{a^1,2}; a^1) = 0.5$, $p(\nu_{a^2,1}; a^2) = 0.3$ and $p(\nu_{a^2,2}; a^2) = 0.7$. In more plain terms, for each token emanating from x^0 through a^1, either one copy is generated at leaf node x^1 and two copies at leaf node x^2 with probability 0.5, or a single copy is generated at leaf node x^2 with probability 0.5. On the other hand, for each token emanating from x^0 through a^2, either one copy is generated at leaf node x^2 with probability 0.3, or one copy is generated at leaf node x^3 with probability 0.7. Finally, the visitation requirement vector is $\mathcal{N} = [2, 1, 1]$. □

The induced MDP problem The ONV problem defined above can be further abstracted to a *Discrete Time Markov Decision Process*, $\mathcal{M} = (S, A, t, c)$, where

- S is the finite set of states, identified with tuples $(\mathcal{X}, \mathcal{N}^c)$. The component $\mathcal{X}$ of this tuple is a vector of dimensionality $|X| - |X^L|$ where each component $\mathcal{X}_x$ denotes the number of tokens at node

$x \in X \setminus X^L$. On the other hand, the component $\mathcal{N}^c$ is a vector belonging in $\prod_{x \in X^L}\{0,\ldots,\mathcal{N}_x\}$ and it expresses the remaining visitation requirements.

- A is a set function defined on S that maps each state $s \in S$ to the finite, non-empty set $A(s)$, comprising all the actions that are feasible in s. More specifically, for $s = (\mathcal{X},\mathcal{N}^c)$ with $\mathcal{X} > \mathbf{0}$, $A(s)$ coincides with $\bigcup_{x \in X \setminus X^L : \mathcal{X}_x > 0} \mathcal{A}(x)$. Furthermore, for all states $s = (\mathcal{X},\mathcal{N}^c)$ with $\mathcal{X} = \mathbf{0}$ and $\mathcal{N}^c \neq \mathbf{0}$, $A(s)$ consists of the single "resetting" action β.

- $t : S \times \bigcup_{s \in S} \mathcal{A}(s) \times S \longrightarrow [0,1]$ is the MDP *state transition* function, i.e., a function on all triplets (s,a,s') with $t(s,a,s')$ being the probability to reach state s' from state s on action a. More specifically, for $s = (\mathcal{X},\mathcal{N}^c)$, $a \in A(s)$ and $s' = (\mathcal{X}',\mathcal{N}^{c'})$,

$$t(s,a,s') = \begin{cases} p(\nu_{a,i};a), & \text{if } a \neq \beta,\ \mathcal{X}'_y = \mathcal{X}_y - 1 \geq 0,\ a \in \mathcal{A}(y),\ \mathcal{X}'_x = \mathcal{X}_x + \nu^x_{a,i},\ \forall x \in X/X^L \\ & \text{with } x \neq y, \text{ and } \mathcal{N}^{c'}_x = (\mathcal{N}^c_x - \nu^x_{a,i})^+,\ \forall x \in X^L,\ 1 \leq i \leq |\mathcal{S}(a)|; \\ 1, & \text{if } a = \beta,\ \mathcal{X} = \mathbf{0},\ \mathcal{X}' = \mathbf{1}^0; \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

  In Equation 1, $\mathbf{1}^0$ denotes the unit vector with all its components equal to zero except for the one corresponding to x^0.

- $c : S \longrightarrow \{0,1\}$ is the *cost function*, where for $s = (\mathcal{X},\mathcal{N}^c)$,

$$c(s) = \begin{cases} 1, & \text{if } \mathcal{X} = \mathbf{0},\ \mathcal{N}^c \neq \mathbf{0}, \\ 0, & \text{if otherwise.} \end{cases} \tag{2}$$

Notice that the cost function defined by Equation 2 assigns a unit cost to every resetting action, but only when there is at least one leaf node with a positive requirement. Hence, the set of states $s = (\mathcal{X},\mathcal{N})$ with $\mathcal{N}^c = \mathbf{0}$ constitute a *closed* class which is also cost-free, i.e., once the process enters this class of states it will remain in it and there will be no more cost accumulation. We shall represent this entire class of states with a single aggregate state, s^T, which we shall refer to as the problem *terminal state*; clearly, s^T is *absorbing* and *cost-free* under any policy π. In order to ensure the reachability of s^T from the initial state s^0, it is further assumed that for every node $x \in X^L$ with $\mathcal{N}_x > 0$, there exists at least one sequence $\xi(x) = a^{(0)}s^{(0)}a^{(1)}s^{(1)}\ldots a^{(k(x))}s^{(k(x))}$ such that (i) $a^{(0)} \in A(s^0)$ with $t(s^0,a^{(0)},s^{(0)}) > 0$, (ii) $\forall i = 1,\ldots,k(x)$, $a^{(i)} \in A(s^{(i-1)})$ with $t(s^{(i-1)},a^{(i)},s^{(i)}) > 0$, and (iii) $s^{k(x)} = (\mathcal{X},\mathcal{N}^c)$ with $\mathcal{N}^c_x < \mathcal{N}_x$; we shall refer to this sequence as an *action path* from node x^0 to node x.

In the following, we are especially interested in a policy π^*, that, starting from the *initial state* $s^0 \equiv (\mathbf{1}^0,\mathcal{N})$, will drive the underlying process to the terminal state s^T with the minimum expected total cost. Let $V_\pi(s^0) = E_\pi[\sum_{t=0}^{\infty} c(s_t)|s_0 = s^0]$, where π is some given policy from the policy set Π, and the expectation $E_\pi[\cdot]$ is taken over all possible realizations under π. Then π^* is formally defined by

$$\pi^* = \arg\min_{\pi \in \Pi} V_\pi(s^0) \tag{3}$$

It is easy to see that, under the aforestated assumptions, the resulting SSP problem is well defined. Therefore, according to [2], there exists a unique vector $V^*(s)$, $s \in S$, with $V^*(s^T) = 0$ and with its remaining components satisfying the Bellman equation

$$V^*(s) = \min_{a \in A(s)} \{c(s) + \sum_{s' \in S} t(s, a, s') \cdot V^*(s')\} \tag{4}$$

The vector $V^*(s)$ is known as the *optimal value function* or the *optimal cost-to-go vector* for the considered MDP formulation. Each component of $V^*(s)$ expresses the expected total cost of initiating the underlying process at state $s \in S$ and subsequently following an optimal policy. Furthermore, its availability enables the specification of an optimal policy π^*, by setting for all $s \in S \backslash \{s^T\}$,

$$\pi^*(s) := \arg \min_{a \in A(s)} \{c(s) + \sum_{s' \in S} t(s, a, s') \cdot V^*(s')\} \tag{5}$$

The computational complexity of the considered ONV problem A close consideration of the SSP formulation defined in the previous paragraph will reveal that the size of its state space is $O(\prod_{x \in X^L} \mathcal{N}_x)$, and therefore, its solution through classical MDP methods is an intractable proposition, for most practical cases. In this paragraph we establish that the ONV problem considered in this section is PSPACE-hard [7], and therefore, the aforementioned intractability is an inherent property of the problem and not a deficiency of the applied methodology. More specifically, the next theorem establishes the PSPACE-hardness of the considered ONV problem through a polynomial reduction of the well-known QSAT problem [7] to its *decision version*, which is defined by the following question: Given an ONV problem instance $\mathcal{E}$ and an integer K, is there a policy π with an expected value $V_\pi < K$?

Theorem 1 *The decision version of the ONV problem with "split" transitions is PSPACE-hard.*

Proof: As mentioned above, to show PSPACE-hardness, we reduce QSAT to the considered problem.[4] For any quantified formula ϕ with n variables and m clauses, we construct an ONV problem instance, $\mathcal{E}(X, \mathcal{A}, \mathcal{P}, \mathcal{N}; \phi)$, that involves an acyclic graph with n decision and $m + 1$ terminal nodes, and its optimal policy has a cost of 1 if and only if the original QSAT problem is satisfiable.

We now proceed into the details of the construction (cf. Figure 2 for a concrete example). The acyclic graph consists of n decision nodes, partitioned in n consecutive layers, corresponding to the n variables $x_1, \ldots, x_n$. A decision node corresponding to an existential variable has two emanating decision arcs whereas a decision node corresponding to a universal variable has one. Furthermore, we assume $m + 1$ leaf nodes, with the first m corresponding to the m clauses $c_1, \ldots, c_m$ of the boolean formula ϕ. Each decision arc emanating from an existential node corresponds to a truth assignment of the corresponding variable. Each such decision arc leads with certainty to a multi-set that (i) drives tokens to the leaf nodes corresponding to the satisfied clauses or, if no clause is satisfied, a token to the $(m+1)^{th}$ leaf node, and (ii) drives one more token to the decision node in the subsequent layer. On the other hand, the

[4] We remind the reader that in the QSAT problem we are given a quantified boolean formula with alternating quantifiers, $\exists x_1 \forall x_2 \exists x_3 \ldots \forall x_n, \phi(x_1, \ldots, x_n)$ and we seek to determine whether this formula is *satisfiable*, that is, whether there is a truth value for x_1 such that for all truth values of x_2, etc. there is a truth value of x_n, such that ϕ comes out true.

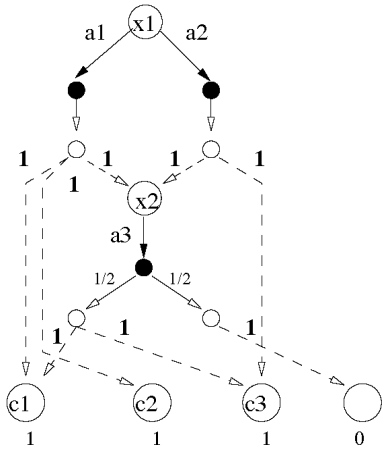

Figure 2: The acyclic graph for the ONV problem that corresponds to the quantified boolean formula $\exists x_1 \forall x_2, \phi(x_1, x_2)$, where $\phi(x_1, x_2)$ is the conjunction of the following three clauses: $c_1 = x_1 \vee x_2, c_2 = x_1$ and $c_3 = \overline{x}_1 \vee x_2$. The dashed lines indicate the multi-sets corresponding to each decision.

single decision arc that corresponds to a universal node leads to two distinct multi-sets with probability $\frac{1}{2}$. Each such multi-set corresponds to a truth assignment for the corresponding universal variable, and is constructed in a similar fashion as before. Finally, we assign a unit requirement to the first m leaf nodes and a requirement of zero to the last leaf node of the acyclic graph.

We claim that the optimal expected cost of $\mathcal{E}(X, \mathcal{A}, \mathcal{P}, \mathcal{N}; \phi)$ is equal to one if and only if formula ϕ is satisfiable. Suppose that the optimal expected cost is 1; i.e., we can choose a decision at the first decision node such that for any multi-set chosen in the second node, there is a decision in the third node etc., such that all leaf nodes satisfy their unit requirement. Then, it is obvious that this policy defines a truth assignment for the first existential variable x_1 such that for every truth assignment of the second variable x_2, there is a truth assignment to x_3 etc., such that all the clauses are satisfied. Conversely, if the quantified formula $\exists x_1 \forall x_2 \exists x_3 \ldots \forall x_n, \phi$ is true, there is a truth assignment for x_1, such that for every truth assignment of x_2, there is a truth assignment for x_3 etc., such that ϕ comes out true. This last statement can be translated into a policy for choosing the appropriate decisions so that at least one token reaches every one of the first m leaf nodes in a single traversal of the corresponding graph, thus resulting in an optimal expected cost of one. □

The "Relaxing LP" and the policy π^{rel} As observed in the previous paragraph, the result of Theorem 1 implies that the exact solution of the considered ONV formulation is an intractable proposition for most problem instances. Hence, we are motivated to seek efficient and computationally tractable suboptimal policies. The policy developed next satisfies these requirements, also being *asymptotically optimal*, since the ratio of its expected performance to V^* converges to unity as the node visitation requirement vector, $\mathcal{N}$, is scaled to infinity. We shall refer to this policy as π^{rel}. Its definition and the aforementioned properties derive from a continuous – or *"fluid"* – relaxation of the considered ONV problem, that is expressed by the following LP formulation:

$$V^*_{rel} \equiv \min \sum_{a \in \mathcal{A}(x^0)} \chi_a \tag{6}$$

s.t.

$$\forall x \in X \setminus (\{x^0\} \cup X^L),$$

$$\sum_{a \in \bigcup_{y \in X \setminus X^L}} \sum_{\mathcal{A}(y)\ 1 \le i \le |\mathcal{S}(a)|} p(\nu_{a,i}; a) \nu^x_{a,i} \chi_a = \sum_{a \in \mathcal{A}(x)} \chi_a \tag{7}$$

$$\forall x \in X^L,$$

$$\sum_{a \in \bigcup_{y \in X \setminus X^L}} \sum_{\mathcal{A}(y)\ 1 \le i \le |\mathcal{S}(a)|} p(\nu_{a,i}; a) \nu^x_{a,i} \chi_a \ge \mathcal{N}_x \tag{8}$$

$$\forall x \in X \setminus X^L,\ \forall a \in \mathcal{A}(x), \quad \chi_a \ge 0 \tag{9}$$

In the following, we shall refer to the above LP formulation as the *"relaxing LP"*. Any optimal solution, χ^*, of the relaxing LP can be naturally interpreted as a *generalized* flow pattern that can satisfy the flow requirements for the terminal nodes $x \in X^L$ expressed by the visitation requirement vector, $\mathcal{N}$, while minimizing the total amount of flow induced into the graph. In particular, the generalized nature of the flow results from the fact that in Equations 7-8 the flow leaving a node, x, is magnified by the gains defined by the multi-sets $\nu_{a,i},\ 1 \le i \le |\mathcal{S}(a)|$. Then, the constraints corresponding to Equation 7 express a "balance" requirement for the generalized flow that is conveyed through the internal nodes of the underlying acyclic digraph, while the constraints corresponding to Equation 8 express the requirement that the total amount of flow conveyed to each terminal node $x \in X^L$ is at least equal to the corresponding visitation requirement $\mathcal{N}_x$ in the original ONV problem.

Example 2 Consider the problem instance described in Example 1 and depicted in Figure 1. The corresponding relaxing LP is expressed by the following linear program:

$$\min\ \chi_{a^1} + \chi_{a^2}$$

s.t.

$$\begin{aligned} 0.5 \cdot 1 \cdot \chi_{a^1} &\ge 2 \\ 0.5 \cdot 2 \cdot \chi_{a^1} + 0.5 \cdot 1 \cdot \chi_{a^1} + 0.3 \cdot 1 \cdot \chi_{a^2} &\ge 1 \\ 0.7 \cdot 1 \cdot \chi_{a^2} &\ge 1 \\ \chi_{a^1} \ge 0,\ \chi_{a^2} &\ge 0 \end{aligned}$$

□

Given an optimal solution $\chi^* = \{\chi^*_a |\ a \in \bigcup_{x \in X \setminus X^L} \mathcal{A}(x)\}$ of the LP defined by Equations 6-9, we define the aforementioned randomized policy π^{rel} as follows: π^{rel} assigns to a state $s = (\mathcal{X}, \mathcal{N}^c)$ with $\mathcal{X} \neq \mathbf{0}$ an action $\pi^{rel}(\mathcal{X}, \mathcal{N}^c)$ by (i) randomly picking a node $x \in X \setminus X^L$ with $\mathcal{X}_x > 0$ and (ii) executing

an action $\pi^{rel}(x;\mathcal{X},\mathcal{N}) \in \mathcal{A}(x)$ on a single token according to the probability distribution

$$\text{Prob}(\pi^{rel}(x;\mathcal{X},\mathcal{N}^c) = a) = \frac{\chi^*_a}{\sum_{a\in\mathcal{A}(x)} \chi^*_a},\ a \in \mathcal{A}(x) \tag{10}$$

For states $s = (\mathcal{X},\mathcal{N})$, with $\mathcal{X} = \mathbf{0}$ and $\mathcal{N}^c > 0$, the policy executes with certainty the *"resetting"* action $\beta \in A(s)$ that initiates a new graph traversal.

Example 3 It can be easily verified that the LP of Example 2 has the unique optimal solution $(\chi^*_{a^1}, \chi^*_{a^2}) = (4.00, 1.43)$ (in a two-digit accuracy). The randomized policy π^{rel} that is induced by this solution for the ONV problem instance of Figure 1, will select, at every graph traversal, action a^1 with probability $p(a^1) = 4/(4+1.43) \approx 0.737$ and action a^2 with probability $p(a^2) = 1.43/(4+1.43) \approx 0.263$. □

It should be obvious from the above discussion, that the relaxing LP of Equations 6-9 involves a number of variables and constraints that is polynomially related to the size of the underlying ONV problem. Since it is also true that the solution of an LP formulation is of polynomial complexity with respect to the number of the variables and the constraints involved, it can be concluded that the aforestated policy π^{rel} can be deployed with a polynomial complexity in terms of the ONV problem size $|\mathcal{E}|$. Some further reflection on the specification of policy π^{rel} will also reveal that an element which is instrumental for the establishment of its polynomial complexity is the fact that the policy maintains the action selection probabilities fixed during every traversal of the underlying stochastic digraph, essentially ignoring the information provided by the vector of the remaining visitation requirements, $\mathcal{N}^c$. We shall refer to the class of randomized policies for the considered ONV problem that are characterized by such an invariance to the vector of the remaining visitation requirements, $\mathcal{N}^c$, as *static* randomized policies. Furthermore, in the following, the space of static randomized policies will be denoted by Π^S and the optimal value of any given ONV problem instance restricted in Π^S will be denoted by V^*_S. In the rest of this section we establish that, in spite of its static nature, policy π^{rel} is an asymptotically optimal policy for the ONV problem with respect to the broader policy set Π. In order, however, to develop this result, it is necessary first to introduce some additional properties of the relaxing LP, including its ability to provide a lower bound for the optimal value, V^*, of its originating ONV problem.

The optimal value of the relaxing LP as a lower bound to V^* Consider an optimal solution to the relaxing LP, χ^*, and let e^{rel}_j denote the amount of flow reaching leaf node x^j when a unit amount of flow is induced into the graph and it is conveyed according to the flow pattern defined by the routing probabilities of Eq. 10. Then, the linearity of Constraint 7 implies that e^{rel}_j can be formally expressed by the equation

$$e^{rel}_j = \frac{\sum_{a\in\bigcup_{y\in X/X^L}\mathcal{A}(y)} \sum_{1\le i\le|\mathcal{S}(a)|} p(\nu_{a,i};a)\nu^x_{a,i}\chi_a}{\sum_{a\in\mathcal{A}(x^0)}\chi^*_a} \tag{11}$$

i.e., as the ratio of the total flow conveyed to the terminal node x_j by the optimal solution χ^* to the total flow V^*_{rel} that is conveyed by χ^* through the entire acyclic digraph. Also, the same property when

combined with the above definition of e_j^{rel} further implies that

$$V_{rel}^* = \max_{j:\mathcal{N}_j>0} \{\frac{\mathcal{N}_j}{e_j^{rel}}\} \tag{12}$$

The quantities e_j^{rel} admit also a natural interpretation in the original ONV problem context. More specifically, a basic inductive argument on the number of layers of the node set X can establish that e_j^{rel} is equal to the *expected* number of tokens reaching leaf node x^j during a single graph traversal under the policy π^{rel} that is induced by χ^*. Finally, an argument similar to that provided in the proof of Theorem 3 in [1][5] can further establish that

$$V_{rel}^* \leq V^* \tag{13}$$

We formalize the above two results of Equations 12 and 13 in the following theorem:

Theorem 2 *Given a problem instance $\mathcal{E} = (X, \mathcal{A}, \mathcal{P}, \mathcal{N})$, let V_{rel}^* and χ^* respectively denote the optimal value and an optimal solution of the relaxing LP. Also, let e_j^{rel}, $x^j \in X^L$, be defined from χ^* according to Equation 11. Then,*

$$V_{rel}^* = \max_{j:\mathcal{N}_j>0} \{\frac{\mathcal{N}_j}{e_j^{rel}}\} \leq V^* \tag{14}$$

Establishing the asymptotic optimality of π^{rel} Before we proceed with the main developments of this paragraph, we present a technical lemma that is necessary in the subsequent derivations. The proof of this lemma is based on results coming from renewal theory and it can be found in the Appendix.

Lemma 1 *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables such that $\forall i,\ 0 \leq X_i \leq K$ almost surely, and $E[X_1] = \mu$. Set $S_0 = 0$; $S_k = \sum_{i=1}^{k} X_i$, $\forall k \geq 1$, and define $\psi_n = \max\{k : S_k \leq n \cdot c\}$, $\forall n \geq 1$. Then the sequence of random variables*

$$\{n^{-r/2}(\psi_n - \frac{n \cdot c}{\mu})^r, n \geq 1\} \tag{15}$$

is uniformly integrable for every $r \geq 1$.

Next we proceed to prove the asymptotic optimality of π^{rel}. For this, consider the problem sequence, $\{\mathcal{E}(n)\}$, that is induced by a problem instance $\mathcal{E} = (X, \mathcal{A}, \mathcal{P}, \mathcal{N})$ through the scaling of the visitation requirement vector, $\mathcal{N}$, by a factor $n \in \mathbb{Z}^+$. Also, in the following, we shall let $\{V_{rel}^*(n)\}$ denote the sequence of the optimal objective values of the relaxing LP implied by the problem sequence $\{\mathcal{E}(n)\}$, and $\{V^*(n)\}$ denote the sequence of the corresponding optimal expected total costs. It is important to notice that, as we scale the requirement vector, $\mathcal{N}$, by a factor n, the optimal solutions, χ^*, of the relaxing-LP are scaled by the same factor $n \in \mathbb{Z}^+$. More specifically, we have the following lemma:

[5]The gist of this argument is as follows: Consider the *"dual LP"* [2] of the MDP that corresponds to the SSP formulation of the considered ONV problem. Then, any feasible solution of this formulation admits a flow interpretation on the state space of the ONV problem [2]. Furthermore, the aggregation of this flow, that traverses the state space of the ONV problem, across the arcs of the underlying state transition diagram that correspond to the same transitions in the problem defining graph $\mathcal{G}$, will provide another flow that constitutes a feasible solution to the relaxing LP. In addition, the original and the induced flows result in the same objective values for their corresponding formulations. But then, it is clear that the relaxing LP is indeed a relaxation of the original ONV formulation and Equation 13 follows from this result.

Lemma 2 *Let $\chi^*(n)$ denote an optimal solution of the relaxing-LP that is obtained through the uniform scaling of the visitation requirement vector $\mathcal{N}$ by a factor $n \in \mathbb{Z}^+$. Then,*

$$\chi^*(n) = n \cdot \chi^*(1) \tag{16}$$

where $\chi^*(1)$ denotes an optimal solution for the relaxing LP that corresponds to $n = 1$, i.e., the original ONV problem instance, and

$$V^*_{rel}(n) = n \cdot V^*_{rel}(1) \tag{17}$$

Proof: Assume that $\mathbf{B}$ is the matrix of an *optimal basis* [15] for the relaxing-LP expressed by Equations 6-9. Then, the corresponding optimal solution of the relaxing-LP is obtained by the vector of the *basic variables* $\mathbf{x}_B = \mathbf{B}^{-1}\mathcal{N}$ while setting the non-basic variables equal to zero. The replacement of the right hand side vector $\mathcal{N}$ in the constraints of Equation 8 by the scaled vector $n \cdot \mathcal{N},\ n \in \mathbb{Z}^+$, preserves the optimality of basis $\mathbf{B}$ (Chapter 5 of [15]), and the new vector of the *basic variables*, $\mathbf{x}_B(n)$, is given by

$$\begin{aligned} \mathbf{x}_B(n) &= \mathbf{B}^{-1}(n \cdot \mathcal{N}) \\ &= n \cdot \mathbf{B}^{-1}\mathcal{N} \\ &= n \cdot \mathbf{x}_B \quad = \quad n \cdot \mathbf{x}_B(1),\ n \in \mathbb{Z}^+ \end{aligned} \tag{18}$$

But then, Equations 16, 17 follow from Equation 18 and the definition of $\mathbf{x}_B(n)$ and $\mathbf{x}_B$. □

When combined with Equation 10, the results of Lemma 2 imply that the set of policies $\pi^{rel}(n)$ that are obtained for the ONV problem instances $\mathcal{E}(n)$ according to the process delineated in the previous paragraphs, is invariant with respect to n. In other words, every policy $\pi^{rel}(n)$ that is obtained for the problem instance $\mathcal{E}(n)$ through an optimal solution $\chi^*(n)$ of the corresponding relaxing LP, is also one of the policies π^{rel} that are obtained for the original ONV problem instance $\mathcal{E}$. Hence, for the rest of this paper, we shall refer to $\pi^{rel}(n)$ as π^{rel}, and we shall define $\{V^{\pi^{rel}}(n)\}$ as the sequence of the expected costs incurred to the problem instances $\mathcal{E}(n)$ by the application of a given instantiation of the randomized policy π^{rel}, that is obtained through Equation 10 and an optimal solution χ^* of the relaxing LP of Equations 6–9. Then, we have the following theorem:

Theorem 3 *Given a problem instance $\mathcal{E} = (X, \mathcal{A}, \mathcal{P}, \mathcal{N})$, consider the problem sequence $\mathcal{E}(n)$ that is obtained through the uniform scaling of the visitation requirement vector $\mathcal{N}$ by a factor $n \in \mathbb{Z}^+$. Then, as $n \to \infty$,*[6]

$$V^{\pi^{rel}}(n) - V^*_{rel}(n) = O(\sqrt{n}) \tag{19}$$

Furthermore, if there exists a target leaf node x^k such that, for any other target leaf node x^j, $\frac{\mathcal{N}_k}{e^{rel}_k} > \max_{j \neq k}\{\frac{\mathcal{N}_j}{e^{rel}_j}\}$, then, as $n \to \infty$,

$$V^{\pi^{rel}}(n) - V^*_{rel}(n) = O(1) \tag{20}$$

[6] We remind the reader that $f(n) = O(g(n)) \Rightarrow \exists c, n_0$ s.t. $0 \leq f(n) \leq c \cdot g(n),\ \forall n \geq n_0$.

Proof: Consider the problem instance $\mathcal{E}(n)$, and let X_j^i denote the random number of tokens ending at leaf node x^j during the i^{th} graph traversal under policy π^{rel}. Then, $\{X_j^i :\ i = 1, 2, \ldots\}$ are sequences of non-negative, identically distributed and independent random variables with $E[X_j^i] = e_j^{rel}$. The quantities e_j^{rel} are defined by Equation 11 on the basis of the optimal solution of the relaxing LP, χ^*, that was employed in the specification of the policy π^{rel}. Furthermore, we define $\sigma_j^2 \equiv Var(X_j^i)$, and we notice that these variances are finite, since the random variables X_j^i have finite support. Finally, we define the *renewal sequence* $S_j^k \equiv \sum_{i=1}^k X_j^i$ and we let $\{\psi_j^n,\ n \geq 0\}$ be a *renewal process* [6] associated with the sequence $\{S_j^k\}$. Hence, for every $j : \mathcal{N}_j > 0$, ψ_j^n is defined by

$$\psi_j^n = \max\{k :\ S_j^k \leq n \cdot \mathcal{N}_j\} \tag{21}$$

and with the additional convention that $\psi_j^n = 0$ if $X_j^1 > n \cdot \mathcal{N}_j$. It is evident from the above definitions that the sequence S_j^k denotes the number of tokens ending at terminal node x^j during the first k graph traversals. Therefore, $1 + \psi_j^n$ is an upper bound to the number of graph traversals necessary to cover the requirements $n \cdot \mathcal{N}_j$ at node x^j. Hence, an upper bound on the total number of graph traversals necessary to cover the total number of requirements, $n \cdot |\mathcal{N}|$, is given by $\max_{j:\mathcal{N}_j>0}\{1 + \psi_j^n\}$. Consequently, the performance of policy π^{rel} on $\mathcal{E}(n)$ satisfies

$$V^{\pi^{rel}}(n) \leq E[\max_{j:\mathcal{N}_j>0}\{1 + \psi_j^n\}] \tag{22}$$

Furthermore, from Lemma 2 we have that $V_{rel}^*(n) = nV_{rel}^*(1)$, which when combined with Theorem 2, imply that

$$V_{rel}^*(n) = \max_{j:\mathcal{N}_j>0}\{\frac{n\mathcal{N}_j}{e_j^{rel}}\} \tag{23}$$

Therefore,

$$\begin{aligned} V^{\pi^{rel}}(n) - V_{rel}^*(n) &\leq 1 + E[\max_{j:\mathcal{N}_j>0}\{\psi_j^n\}] - \max_{j:\mathcal{N}_j>0}\{\frac{n\mathcal{N}_j}{e_j^{rel}}\} \\ &\leq 1 + E[\max_{j:\mathcal{N}_j>0}\{|\psi_j^n - \frac{n\mathcal{N}_j}{e_j^{rel}}|\}] \\ &\leq 1 + \sum_{j:\mathcal{N}_j>0} E[|\psi_j^n - \frac{n\mathcal{N}_j}{e_j^{rel}}|] \end{aligned} \tag{24}$$

where the first inequality is the result of Equations 22-23 and the second inequality is the result of the following property:

$$\begin{aligned} &\forall a_i, b_i \in \mathbb{R}, i = 1, \ldots, n, \\ |\max\{a_1, a_2, \ldots, a_n\} - \max\{b_1, b_2, \ldots, b_n\}| &\leq \max\{|a_1 - b_1|, |a_2 - b_2|, \ldots, |a_n - b_n|\} \end{aligned}$$

Also, from the *renewal central limit theorem* [6] we get that

$$\frac{1}{\sqrt{n}} \cdot (\psi_j^n - \frac{n\mathcal{N}_j}{e_j^{rel}}) \Rightarrow N(0, \frac{\sigma_j^2 \cdot \mathcal{N}_j}{(e_j^{rel})^3}),\ j : \mathcal{N}_j > 0 \tag{25}$$

as $n \to \infty$. But then, Equation 25, when combined with Lemma 1 and the Continuous Mapping Theorem, imply that, for $r > 0$,

$$(\frac{1}{\sqrt{n}})^r E[|\psi_j^n - \frac{n\mathcal{N}_j}{e_j^{rel}}|^r] \longrightarrow E[|N(0, \frac{\sigma_j^2 \cdot \mathcal{N}_j}{(e_j^{rel})^3})|^r],\ j : \mathcal{N}_j > 0 \tag{26}$$

as $n \to \infty$. Equation 19 now follows from Equation 24 when combined with Equation 26.

To prove Equation 20 we proceed as follows: Assume that $\max_{j:\mathcal{N}_j>0}\{\frac{n\mathcal{N}_j}{e_j^{rel}}\} = \frac{n\cdot\mathcal{N}_1}{e_1^{rel}}$; then,

$$\begin{aligned} V^{\pi^{rel}}(n) - V^*_{rel}(n) &\leq 1 + E[\max_{j:\mathcal{N}_j>0}\{\psi_j^n\}] - \max_{j:\mathcal{N}_j>0}\{\frac{n\mathcal{N}_j}{e_j^{rel}}\} \\ &= 1 + E[\max_{j:\mathcal{N}_j>0}\{\psi_j^n\}] - E[\psi_1^n] + E[\psi_1^n] - \frac{n\mathcal{N}_1}{e_1^{rel}} \\ &= 1 + E[\max_{j:\mathcal{N}_j>0}\{\psi_j^n - \psi_1^n\}] + E[\psi_1^n] - \frac{n\mathcal{N}_1}{e_1^{rel}} \\ &\leq 1 + \sum_{j\neq 1:\mathcal{N}_j>0} E[(\psi_j^n - \psi_1^n)^+] + E[\psi_1^n] - \frac{n\mathcal{N}_1}{e_1^{rel}} \end{aligned} \tag{27}$$

Since, for every $n \geq 1$, $\psi_j^n + 1$ is a *stopping time* with respect to $\{X_j^i\}$, with $E[\psi_j^n] < \infty$, we can write [6]

$$\begin{aligned} E[\sum_{i=1}^{\psi_j^n+1} X_j^i] &= E[\psi_j^n + 1]E[X_j^1] \\ &= e_j^{rel}\cdot(E[\psi_j^n] + 1) \end{aligned} \tag{28}$$

Let K denote the maximum number of tokens that can be generated during a single graph traversal. Then, by definition of $\psi_j^n + 1$,

$$n\cdot\mathcal{N}_j \leq \sum_{i=1}^{\psi_j^n+1} X_j^i \leq n\cdot\mathcal{N}_j + K,\ j:\mathcal{N}_j>0. \tag{29}$$

Equations 28 and 29 imply that

$$0 \leq E[\psi_j^n] + 1 - \frac{n\cdot\mathcal{N}_j}{e_j^{rel}} \leq \frac{K}{e_j^{rel}} \tag{30}$$

Next, we prove that

$$E[(\psi_j^n - \psi_1^n)^+] \to 0,\ \forall j:\mathcal{N}_j>0 \tag{31}$$

as $n \to \infty$. Indeed, for $r \geq 1$, $a_j^n = \frac{1}{\sqrt{n}}(\psi_j^n - \frac{n\cdot\mathcal{N}_j}{e_j^{rel}})$ and $c_j = \frac{\mathcal{N}_1}{e_1^{rel}} - \frac{\mathcal{N}_j}{e_j^{rel}} > 0$, we have that

$$\begin{aligned} E[(\psi_j^n - \psi_1^n)^+] &= E[(\psi_j^n - \psi_1^n)\cdot I(\psi_j^n \geq \psi_1^n)] \\ &\leq E[\psi_j^n \cdot I(\psi_j^n \geq \psi_1^n)] \\ &\leq [E[(\psi_j^n)^2]\cdot P(\psi_j^n \geq \psi_1^n)]^{1/2} \\ &= [E[(\psi_j^n)^2]\cdot P((\psi_j^n - \frac{n\cdot\mathcal{N}_j}{e_j^{rel}}) \\ &- (\psi_1^n - \frac{n\cdot\mathcal{N}_1}{e_1^{rel}}) \geq \frac{n\cdot\mathcal{N}_1}{e_1^{rel}} - \frac{n\cdot\mathcal{N}_j}{e_j^{rel}})]^{1/2} \\ &= [E[(\psi_j^n)^2]\cdot P(a_j^n - a_1^n \geq \sqrt{n}\cdot c_j)]^{1/2} \\ &\leq [E[(\psi_j^n)^2]\cdot \frac{1}{c_j^r\cdot n^{r/2}}\cdot E[(a_j^n - a_1^n)^r]]^{1/2} \\ &\leq [E[(\psi_j^n)^2]\cdot \frac{2^{r-1}}{c_j^r\cdot n^{r/2}}\cdot E[|a_j^n|^r + |a_1^n|^r]]^{1/2} \end{aligned} \tag{32}$$

where the second inequality is an application of Schwarz inequality, the third inequality is an application of Markov inequality, and the last inequality is a direct consequence of $(a+b)^r \leq 2^{r-1} \cdot (|a|^r + |b|^r),\ a, b \in R$. Furthermore, from Theorem 2.3 of [16] we have that

$$E[(\psi_j^n)^2] = O(n^2). \tag{33}$$

Furthermore, from Equation 26 we have

$$E[|a_j^n|^r + |a_1^n|^r]] \rightarrow E[|N(0, \frac{\sigma_j^2 \cdot \mathcal{N}_j}{(e_j^{rel})^3})|^r + |N(0, \frac{\sigma_1^2 \cdot \mathcal{N}_1}{(e_1^{rel})^3})|^r]]. \tag{34}$$

as $n \rightarrow \infty$. Therefore, from Equations 32, 33 and 34 we get

$$E[(\psi_j^n - \psi_1^n)^+] = O(n^{2-r/2}). \tag{35}$$

Consequently, if we choose r such that $\frac{r}{2} > 2$, Equation 31 holds. Finally, Equation 20 follows immediately from Equation 27 when combined with Equations 30 and 31. $\square$

The asymptotic optimality of π^{rel} is an immediate implication of Theorem 3. This result is formally stated and proven in the following corollary:

Corollary 1 *Given a problem instance $\mathcal{E} = (X, \mathcal{A}, \mathcal{P}, \mathcal{N})$, consider the problem sequence $\mathcal{E}(n)$ that is obtained through the uniform scaling of the visitation requirement vector $\mathcal{N}$ by a factor $n \in \mathbb{Z}^+$. Then, as $n \rightarrow \infty$,*

$$\frac{V^{\pi^{rel}}(n)}{V^*(n)} \longrightarrow 1 \tag{36}$$

Proof: The definitions of $V^{\pi^{rel}}(n)$ and $V^*(n)$ imply that

$$\frac{V^{\pi^{rel}}(n)}{V^*(n)} \geq 1,\ n \in \mathbb{Z}^+. \tag{37}$$

From Lemma 2 we also get that $V_{rel}^*(n) = n \cdot V_{rel}^*(1),\ n \in \mathbb{Z}^+$, and, therefore, Theorem 3 implies that

$$\begin{aligned} \frac{V^{\pi^{rel}}(n)}{V_{rel}^*(n)} &= \frac{V_{rel}^*(n) + O(\sqrt{(}n))}{V_{rel}^*(n)} \\ &= 1 + \frac{O(\sqrt{(}n))}{n \cdot V_{rel}^*(1)} \\ &\rightarrow 1, \quad \text{as} \quad n \rightarrow \infty \end{aligned} \tag{38}$$

Since, from Theorem 2, $V_{rel}^*(n) \leq V^*(n)$, we also have that

$$\frac{V^{\pi^{rel}}(n)}{V^*(n)} \leq \frac{V^{\pi^{rel}}(n)}{V_{rel}^*(n)},\ n \in \mathbb{Z}^+. \tag{39}$$

The corollary follows by combining Equations 37, 38 and 39. $\square$

Next we draw the reader's attention to the second result of Theorem 3, which is expressed by Equation 20. In plain terms, this result implies that when the maximizer of the ratios $\mathcal{N}_j / e_j^{rel}$ is unique, the difference of the expected performance of policy π^{rel} from the lower bound $V_{rel}^*(n)$ to the optimal value $V^*(n)$ remains bounded as n grows to infinity. In particular, this bound is established by

Equations 27, 30 and 31 as K/e_k^{rel}, where x^k is the unique maximizer leaf node of the ratios $\mathcal{N}_j/e_j^{rel}$. In general, results of this type imply an excellent asymptotic performance for the corresponding policy and they are rather scarce in the relevant literature. The considered result is even more surprising when noticing the static nature of policy π^{rel} that was discussed in the previous paragraphs. An apparent intuitive interpretation of it is that the uniqueness of the maximizer of the ratios $\mathcal{N}_j/e_j^{rel}$ defines very prominently a *"most difficult"* target leaf node x^k, to the extent that the bias of policy π^{rel} towards this node[7] remains valid for all but a finite number of task iterations in each problem instance $\mathcal{E}(n)$, as n grows to infinity.

Closing the discussion of this section, we also want to point out that the asymptotic regime involved in the results of Theorem 3 and Corollary 1 is particularly relevant to the ONV formulations that arise in the context of the sampling processes discussed in the introductory section. It is well known that the learning algorithms considered in that section require extensive amounts of sampling in order to deliver the typically sought performance. In particular, the scaling process of the visitation requirements that underlies the asymptotic results presented in this section, is materialized in the context of our learning algorithms by setting their performance parameters ϵ and δ to values increasingly closer to zero. In the next section, we seek to increase the relevance of the developed results to the motivating learning algorithms, by addressing the ONV problem with additional visitation requirements for the non-terminal nodes of the problem-defining graph.

3 Adding the Internal Visitation Requirements

The new ONV problem version In this section we consider the extension of the ONV problem addressed in Section 2, that is obtained by the introduction of visitation requirements for the internal nodes of the stochastic digraph that underlies the problem definition. An instance of this new ONV problem is defined again by a quadruple $\mathcal{E} = (X, \mathcal{A}, \mathcal{P}, \mathcal{N})$, where all the components remain the same as in the case of Section 2, except for the visitation requirement vector $\mathcal{N}$, which now is defined as follows:

- $\mathcal{N}$ associates with each node $x \in X$ a visitation requirement $\mathcal{N}_x \in \mathbb{Z}_0^+$. The support $||\mathcal{N}||$ of $\mathcal{N}$ is defined by the nodes $x \in X$ with $\mathcal{N}_x > 0$. Furthermore, it is implicitly assumed that the visitation requirements of a node $x \in X$ will start to be satisfied only after the complete satisfaction of the visitation requirements of all its successor nodes.

The new problem described above can be further abstracted to an MDP, $\mathcal{M} = (S, A, t, c)$, where all the components remain the same as in the MDP definition of the ONV problem addressed in Section 2, except for the remaining visitation requirement vector $\mathcal{N}^c$ and its updating through the transition function t. More specifically, in this new problem context, $\mathcal{N}^c$ is an $|X|$-dimensional vector initialized at $\mathcal{N}$. Furthermore, given a state $s = (\mathcal{X}, \mathcal{N}^c) \in S$ with $\mathcal{X}_y > 0$, and a decision $a \in A(y)$, we compute the state $s' = (\mathcal{X}', \mathcal{N}^{c'})$, that results from the execution of a in s through its outcome defined by the multi-set $\nu_{a,i}$, according to the following procedure:

[7] This bias is established during the policy construction by the structure of the employed optimal solution χ^* of the relaxing LP.

1. $\mathcal{X}'_y := \mathcal{X}_y - 1$;

2. $\forall x \in X \backslash X^L, \;\; \mathcal{X}'_x := \mathcal{X}_x + \nu^x_{a,i}$;

3. $\forall l = L, L-1, \ldots, 0, \;\forall x \in X^l$,
   `if` $\sum_{q \in Succ(x)} \mathcal{N}^c_q = 0$ `then` $\mathcal{N}^{c'}_x := (\mathcal{N}^c_x - \nu^x_{a,i})^+$ `else` $\mathcal{N}^{c'}_x := \mathcal{N}^c_x$;

The notation $Succ(x)$ appearing in the above specification denotes the *immediate* successors of node x in the problem-defining graph $\mathcal{G}$.[8] For states $s = (\mathcal{X}, \mathcal{N}^c) \in S$ with $\mathcal{X} = 0$, the process "resets" itself in the spirit expressed by Equation 1 in Section 2. Finally, defining the cost function $c(s)$ and the terminal state s^T as discussed in Section 2, and expressing the problem objective by

$$\pi^* = \arg\min_{\pi \in \Pi} E_\pi[\sum_{t=0}^{\infty} c(s_t) | s_0 = s^0] \tag{40}$$

we obtain a well-defined SSP problem. In the following, we shall use the notation $V^*(s)$ and $\pi^*(s)$, $s \in S \backslash \{s^T\}$, in order to characterize the optimal value function and an optimal policy for this SSP.

Complexity considerations Since the ONV problem variation defined in the previous paragraph subsumes the ONV problem version defined in Section 2, it is clear that it is PSPACE-hard. On the other hand, one can envision ONV problem instances with internal visitation requirements but without any transition "splits". Currently, we lack a clear-cut result regarding the complexity of this last variation of the ONV problem, and the same is true for the complexity of the original ONV problem studied in [1]. However, as an intermediary step towards the characterization of the complexity of the ONV problem with internal visitation requirements, and corroboration for its hard nature, we have managed to show that the well known problem of *"Poisson-tree" scheduling* [7] reduces polynomially to it. Beyond assisting with positioning the ONV problem with internal visitation requirements in the broader landscape of the computational complexity theory, the provided reduction also reveals the affinity of the ONV problem to the problems addressed by the more classical stochastic scheduling theory.

A brief description of the "Poisson-tree" scheduling problem is as follows [7]: The problem is defined by a triplet $\Theta = (m, \tau, \Gamma)$, where

- m denotes the number of the identical processors that are available in the system.
- $\tau = \{T_1, \ldots, T_n\}$ denotes a finite set of tasks that must be processed by the system processors. It is further assumed that the processing time of each task is exponentially distributed with rate equal to one.
- $\Gamma = (\tau, \Omega)$ is a *rooted in-tree* – i.e., a directed acyclic graph with out-degree of at most one – that expresses a set of *precedence constraints* imposed on the task set τ.

The *memoryless property* possessed by the exponential distribution [6] implies that (i) the natural decision epochs for this scheduling problem are determined by the task completion times, and that (ii) the uncompleted tasks can be scheduled preemptively at those points. The interval between two consecutive

[8] Obviously, for nodes $x \in X^L$, $Succ(x) = \emptyset$ and the condition in the "if" statement of item (3) is immediately satisfied.

decision epochs is referred to as a *processing cycle*. The uniformly unit-valued task processing rates imply that (i) a processing cycle involving k processors has an expected duration of $1/k$, and that (ii) the probability for any of the k processed tasks to finish first is also equal to $1/k$. The problem objective is to identify a *schedule* – i.e., a policy for assigning tasks to the available processors at the end of each processing cycle – that respects the imposed precedence constraints and minimizes the expected *makespan* – i.e., the expected completion time of the last task.

While the complexity of a "Poisson-tree" scheduling problem with two processors is polynomial, the complexity of a three-processor version of the problem is an open issue [7]. The next theorem establishes that the ONV problem with internal visitation requirements is at least as difficult as the three-processor "Poisson-tree" scheduling problem.

Theorem 4 *The decision version of the 3-processor "Poisson-tree" scheduling problem reduces polynomially to the decision version of the ONV problem with internal visitation requirements.*

Proof: Given an instance $\Theta = (m, \tau, \Gamma)$ of the "Poisson-tree" scheduling problem, the corresponding instance $\mathcal{E}(\Theta) = (X, \mathcal{A}, \mathcal{P}, \mathcal{N})$ of the ONV problem is defined as follows (the reader is referred to Figure 3 and Table 3 for a more concrete example of this construction):

- $X = \tau \cup \{x^0, x^\lambda\}$. In the graph $\mathcal{G}$ of the constructed ONV problem, x^0 will play the role of the root node, while x^λ is a terminal node with zero requirements that will enable the modeling of the losses resulting from the under-utilization of the system processors.
- The action set $\mathcal{A}$ is defined as follows:
  - For each node $T_i \in \tau$, the action set $\mathcal{A}(T_i)$ is defined by the set of its incoming arcs in graph Γ.
  - The actions set $\mathcal{A}(x^0)$ is defined by all the single, two and three-element subsets of the task set τ, which do not contain pairs of tasks associated through the precedence relationship defined by Γ.
  - Finally, $\mathcal{A}(x^\lambda) = \emptyset$ (as already mentioned, x^λ is a terminal node).
- The transition function $\mathcal{P}$ establishes the following connectivity:
  - For each node $T_i \in \tau$, the action corresponding to an incoming arc (T_j, T_i) leads deterministically to node T_j.
  - The action at node x^0 corresponding to a task set $\{T_i\}$ leads to node T_i with probability $1/3$, and to node x^λ with probability $2/3$. On the other hand, the action corresponding to a task set $\{T_i, T_j\}$ leads to each of these two nodes with respective probability $1/3$, and to node x^λ with the remaining probability. Finally, an action corresponding to a triplet $\{T_i, T_j, T_k\}$ leads to each of these three nodes with respective probability $1/3$.
- The visitation requirement vector $\mathcal{N}$ assigns a *unit* visitation requirement to each node $T_i \in \tau$ and a zero visitation requirement to x^0 and x^λ.

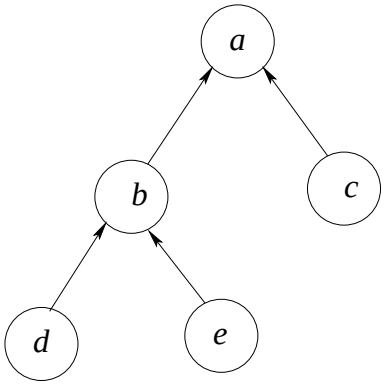

Figure 3: The rooted in-tree modeling the precedence constraints for the tasks of the "Poisson-tree" scheduling problem Θ considered in this example.

Table 1: A tabular characterization of the stochastic graph $\mathcal{G}$ and the visitation requirement vector $\mathcal{N}$ corresponding to the ONV problem instance $\mathcal{E}(\Theta)$.

Node	Action	Outcomes and their Distribution	Visitation Req.
x^0	a_1	$(a, 1/3), (x^l, 2/3)$	0
	a_2	$(b, 1/3), (x^l, 2/3)$	
	a_3	$(c, 1/3), (x^l, 2/3)$	
	a_4	$(d, 1/3), (x^l, 2/3)$	
	a_5	$(e, 1/3), (x^l, 2/3)$	
	a_6	$(b, 1/3), (c, 1/3), (x^l, 1/3)$	
	a_7	$(c, 1/3), (d, 1/3), (x^l, 1/3)$	
	a_8	$(c, 1/3), (e, 1/3), (x^l, 1/3)$	
	a_9	$(d, 1/3), (e, 1/3), (x^l, 1/3)$	
	a_{10}	$(c, 1/3), (d, 1/3), (e, 1/3)$	
a	a_{11}	$(b, 1)$	1
	a_{12}	$(c, 1)$	
b	a_{13}	$(d, 1)$	1
	a_{14}	$(e, 1)$	
c	$\emptyset$		1
d	$\emptyset$		1
e	$\emptyset$		1
x^l	$\emptyset$		0

Clearly, the above construction of $\mathcal{E}(\Theta)$ can be performed in polynomial time with respect to the size of the defining elements of problem Θ. Furthermore, a scheduling decision d applied during a processing cycle of the original problem Θ, can be simulated in the context of the ONV problem $\mathcal{E}(\Theta)$ through the selection of the action $a \in \mathcal{A}(x^0)$ that corresponds to the tasks selected by d, and the resulting outcomes will have the same transition structure in each problem context. At the same time, the deterministic[9] policies applied during any single traversal of the graph $\mathcal{G}$ in problem $\mathcal{E}(\Theta)$ also have a mapping decision in the original problem Θ, with the same transition structure for the resulting outcomes. More specifically, given a state $(x^0, \mathcal{N}^c)$ for problem $\mathcal{E}(\Theta)$, the application over a single traversal of the graph $\mathcal{G}$ of a policy π that, starting from node x^0, selects the action corresponding to a single task T_i and once in the subtree emanating from node T_i follows deterministically a path leading to an active target node T_j, can be interpreted as the scheduling decision of processing only the available task T_j during the corresponding processing cycle of problem Θ. Also, similar interpretations apply to policies π that select actions at state x^0 corresponding to two or three tasks, and subsequently, they reach deterministically one of the target nodes in the resulting subtree. Hence, it is possible to simulate any policy π of Θ on $\mathcal{E}(\Theta)$ and vice versa.

To conclude the proof, it suffices to show that the value functions for any pair of policies $\pi,\ \pi'$ related through the aforementioned simulation, satisfy $V^{\pi}/V^{\pi'} = a$, for some pre-determined constant a (since, then, there will exist a policy π for Θ with $V^{\pi} < K$ *iff* there exists a policy π' for $\mathcal{E}(\Theta)$ with $V^{\pi'} < K/a$). Next we show, through an induction on $|\tau|$, that $a = 1/3$. Indeed, for the base case of $|\tau| = 1$, there will be only one busy processor during the relevant processing cycle, and therefore, $V^{\pi} = 1$, while the simulation of the corresponding decision in the $\mathcal{E}(\Theta)$ context will result in $V^{\pi'} = 3$. For a problem Θ with $|\tau| > 1$, consider that the aforestated relationship holds true for all "Poisson-tree" scheduling problems involving a number of tasks less than or equal to $|\tau| - 1$. Furthermore, let τ^1 denote the set of tasks scheduled by π during the first processing cycle, and also let $\Theta \backslash T_i$ denote the "Poisson-tree" scheduling problem resulting from Θ through the removal from the task set τ of task $T_i \in \tau^1$. Then, it is easy to see that

$$V^{\pi}(\Theta) = (\text{Expected duration of first processing cycle}) + \frac{1}{|\tau^1|} \sum_{T_i \in \tau^1} V^{\pi}(\Theta \backslash T_i) \tag{41}$$

and a similar equation applies to $V^{\pi'}(\mathcal{E}(\Theta))$, i.e.,

$$V^{\pi'}(\mathcal{E}(\Theta)) = (\text{Expected duration until the first visitation}) + \frac{1}{|\tau^1|} \sum_{T_i \in \tau^1} V^{\pi'}(\mathcal{E}(\Theta \backslash T_i)) \tag{42}$$

The induction hypothesis implies that $V^{\pi}(\Theta \backslash T_i)/V^{\pi'}(\mathcal{E}(\Theta \backslash T_i)) = 1/3$ for every task $T_i \in \tau^1$, and the reader can easily verify that the ratio of the first terms in the right-hand-sides of Equations 41 and 42 is also equal to 1/3. Hence, in this case, $V^{\pi}(\Theta)/V^{\pi'}(\mathcal{E}(\Theta)) = 1/3$, as well. □

Problem restriction As observed in the introductory discussion, the fluid relaxation of the ONV problem with internal visitation requirements corresponds to a hybrid optimal control problem. A

[9]Confining this analysis to the set of deterministic policies is enabled by the relevant MDP/SSP theory that guarantees the existence of a deterministic optimal policy.

detailed study of this optimal control problem is provided in [5], but the practical value of the results derived from that analysis is limited by the non-polynomial complexity of the involved computations. Hence, in the following, we constrain the solution of the considered ONV problem over the class of *static* randomized policies, Π^S, which were introduced in the previous section and are simpler in their characterization and evaluation, and more easily implementable. In a spirit similar to that adopted in Section 2, we define a fluid relaxation and an induced randomized policy for the ONV variation considered in this section. However, the proposed fluid relaxation provides a lower bound for V_S^* only,[10] and the induced randomized policy is asymptotically optimal only for the problem restriction in the policy space Π^S.

A computationally efficient and asymptotically optimal policy for the restricted problem The problem relaxation employed in the subsequent analysis is described by the following mathematical programming (MP) formulation:

$$\min Q_{x^0} \tag{43}$$

s.t.

$$\sum_{a \in \mathcal{A}(x^0)} \chi_a = 1 \tag{44}$$

$$\forall x \in X \setminus (\{x^0\} \cup X^L),$$
$$\sum_{a \in \bigcup_{y \in X \setminus X^L} \mathcal{A}(y)} \sum_{1 \le i \le |\mathcal{S}(a)|} p(\nu_{a,i}; a) \nu_{a,i}^x \chi_a = \sum_{a \in \mathcal{A}(x)} \chi_a \tag{45}$$

$$e_{x^0}^{rel} = 1 \tag{46}$$

$$\forall x \in X \setminus \{x^0\},$$
$$e_x^{rel} = \sum_{a \in \bigcup_{y \in X \setminus X^L} \mathcal{A}(y)} \sum_{1 \le i \le |\mathcal{S}(a)|} p(\nu_{a,i}; a) \nu_{a,i}^x \chi_a \tag{47}$$

$$\forall x \in X \setminus \{x^0\} \text{ with } \mathcal{N}_x > 0, \quad e_x^{rel} > 0 \tag{48}$$

$$\forall x \in X^L, \; Q_x = \frac{\mathcal{N}_x}{e_x^{rel}} \tag{49}$$

$$\forall x \in X \setminus X^L, \; Q_x = \max_{y \in Succ(x)} \{Q_y\} + \frac{\mathcal{N}_x}{e_x^{rel}} \tag{50}$$

$$\forall x \in X \setminus X^L, \; \forall a \in \mathcal{A}(x), \quad \chi_a \ge 0 \tag{51}$$

Variables χ_a in the above formulation denote a generalized flow that is routed through the arcs corresponding to the different actions $a \in \mathcal{A}$, and it conveys a unit of fluid that is induced to the problem-defining graph $\mathcal{G}$ through its root node x^0 (c.f., Constraints 44, 45). In a similar spirit, variables e_x^{rel} denote the amount of fluid reaching each node $x \in X$, for each unit of flow induced in $\mathcal{G}$ through

[10] and not for V^*, which was the case with the fluid relaxation of the ONV problem presented in Section 2

node x^0 (c.f., Constraints 46, 47). Furthermore, Constraint 48 requests that any feasible solution of this formulation has a positive flow to every node x with non-zero visitation requirements. Finally, variables Q_x denote the minimum amount of flow required in order to satisfy the corresponding node visitation requirements, under the routing scheme described by variables χ_a, e_x^{rel}, and the precedence constraints expressed by the underlying graph $\mathcal{G}$ (c.f., Constraints 49, 50). In particular, the right-hand-side of Constraint 50 expresses the fact that the accumulation of the fluid requested at an internal node x will take place only after all the flow that is required for the satisfaction of the requirements of its successor nodes has been conveyed through the graph. From a practical computational standpoint, the solution of the above formulation can be further facilitated by replacing Constraint 50 with the following constraint:

$$\forall x \in X \backslash X^L,\ \forall y \in Succ(x),\quad Q_x \geq Q_y + \frac{\mathcal{N}_x}{e_x^{rel}} \tag{52}$$

The resulting formulation is convex, and it can be easily addressed through standard techniques borrowed from convex optimization [17].

Given an optimal flow, χ^*, for the MP formulation defined by Equations 43-51, the definition and execution of the proposed randomized policy follows exactly the same guidelines described in Section 2 for the definition of the policy π^{rel} from the fluid relaxation of the ONV problem addressed in that section. To emphasize this affinity between the two policies, we shall keep referring to the new policy defined in this section as the policy π^{rel}, while the MP formulation of Equations 43-51 will be called the *relaxing MP*. The following theorem is the counterpart of Theorem 2 for this new problem context:

Theorem 5 *Given an instance $\mathcal{E} = (X, \mathcal{A}, \mathcal{P}, \mathcal{N})$ of the ONV problem with internal visitation requirements, let V_{rel}^* and (χ^*, e^{rel^*}, Q^*) respectively denote the optimal value and an optimal solution of the corresponding relaxing MP. Then,*

$$V_{rel}^* = Q_{x^0}^* \leq V_S^* \tag{53}$$

where V_S^* denotes the optimal solution of the considered problem instance when restricted to the space of static randomized policies.

The first part of Equation 53 in Theorem 5 is an immediate implication of the definition of V_{rel}^* and Equation 43. The second part of this equation can be obtained through an argument similar to that outlined in Footnote 5 for the corresponding result of Theorem 2. Furthermore, the perusal of Equations 49–50 reveals that the set of the optimal flows for the relaxing MP, $\{\chi^*\}$, remains invariant as the requirement vector N is scaled uniformly to infinity. This fact subsequently implies the invariance to this scaling of the set of the induced randomized policies, $\{\pi^{rel}\}$. The next theorem and the accompanying corollary establish the asymptotic optimality of any instantiation of policy π^{rel} in the new problem context.

Theorem 6 *Given an instance $\mathcal{E} = (X, \mathcal{A}, \mathcal{P}, \mathcal{N})$ of the ONV problem with internal visitation requirements, consider the problem sequence, $\mathcal{E}(n)$, that is obtained through the uniform scaling of the visitation requirement vector $\mathcal{N}$ by a factor $n \in \mathbb{Z}^+$. Then, as $n \to \infty$,*

$$V^{\pi^{rel}}(n) - V_{rel}^*(n) = O(\sqrt{n}) \tag{54}$$

Proof: As in the proof of Theorem 2, let X_x^i denote the random number of tokens traversing node $x \in X$ during the i^{th} graph traversal under π^{rel} and $\sigma_x^2 = Var(X_x^i)$. Also, let $\{\psi_x^n, n \geq 1\}$ be the renewal process associated with the sequence $\{X_x^i\}$, defined as

$$\psi_x^n = \max\{k : \sum_{i=1}^{k} X_x^i \leq n \cdot \mathcal{N}_x\} \tag{55}$$

with $\psi_x^n = 0$ if $X_x^1 > n \cdot \mathcal{N}_x,\ x : \mathcal{N}_x > 0$. Finally, define

$$\Psi_x^n = \psi_x^n + 1,\ x \in X^L \tag{56}$$

$$\Psi_x^n = \max_{y \in Succ(x)} \{\Psi_y^n\} + \psi_x^n + 1,\ x \in X \backslash X^L \tag{57}$$

Then, the performance of policy π^{rel} satisfies

$$V^{\pi^{rel}}(n) \leq E[\Psi_{x^0}^n] \tag{58}$$

Equation 58, when combined with Theorem 5, imply that, in order to prove the result of Theorem 6, it suffices to show that

$$\forall x \in X, \quad E[|\Psi_x^n - Q_x^*(n)|] = O(\sqrt{n}) \tag{59}$$

where $Q_x^*(n)$ denotes the optimal value of variable Q_x in the relaxing MP formulated for problem instance $\mathcal{E}(n)$.

We proceed to prove this result through an induction on the number of graph layers, l. The base case, for $l = L$, is immediately obtained from the results in the proof of Theorem 3. Next, we consider an l such that $0 \leq l < L$, and assume that Equation 59 holds for all $x \in \bigcup_{l+1 \leq i \leq L} X^i$. Then, for $x \in X^l$, we have that

$$\begin{aligned} E[|\Psi_x^n - Q_x^*(n)|] &= \\ E[|\max_{y \in Succ(x)} \{\Psi_y^n\} + \psi_x^n + 1 - \max_{y \in Succ(x)} \{Q_y^*(n)\} - \frac{n \cdot \mathcal{N}_x}{e_x}|] &\leq \\ E[\max_{y \in Succ(x)} \{|\Psi_y^n - Q_y^*(n)|\}] + E[|\psi_x^n + 1 - \frac{n \cdot \mathcal{N}_x}{e_x}|] &\leq \\ \sum_{y \in Succ(x)} E[|\Psi_y^n - Q_y^*(n)|] + E[|\psi_x^n + 1 - \frac{n \cdot \mathcal{N}_x}{e_x}|] & \end{aligned} \tag{60}$$

Each term of the summation appearing in Equation 60 is $O(\sqrt{n})$ from the induction hypothesis, while the fact that

$$E[|\psi_x^n + 1 - \frac{n \cdot \mathcal{N}_x}{e_x}|] = O(\sqrt{n}) \tag{61}$$

follows immediately from the results in the proof of Theorem 3. But then, the whole quantity appearing in Equation 60 is $O(\sqrt{n})$, establishing the result of Equation 59, and, through that, the result of the Theorem. □

The next corollary derives from Theorems 5 and 6, and it is the counterpart of Corollary 1 for the ONV problem variation considered in this section, when restricted in the class of static randomized policies Π^S.

Corollary 2 *Consider an instance $\mathcal{E} = (X, \mathcal{A}, \mathcal{P}, \mathcal{N})$ of the ONV problem with internal visitation requirements, restricted in the space of static randomized policies Π^S. Also, consider the problem sequence $\mathcal{E}(n)$ that is obtained through the uniform scaling of the visitation requirement vector $\mathcal{N}$ by a factor $n \in \mathbb{Z}^+$. Then, as $n \to \infty$,*

$$\frac{V^{\pi^{rel}}(n)}{V_S^*(n)} \longrightarrow 1 \tag{62}$$

4 Conclusions

The work presented in this paper (i) extended the past results of [1] on the ONV problem to some new problem variations, (ii) initiated the formal complexity analysis of the resulting problem taxonomy, and (iii) offered new insights and a novel methodological base for the analysis of some computationally tractable and asymptotically optimal policies for the addressed variations. From a more practical standpoint, the developed results are important for the effective usage of the emerging theory on the ONV problem in the applications that motivated it. Indeed, a main line of our future work will seek the integration of the insights and results developed in this paper, in the application context presented in the work of [4]; the reader is referred to [18] for some relevant developments. Another line of our future research will seek the formulation of alternative fluid relaxations for the ONV problem with internal visitation requirements, and the investigation of their potential for defining efficient suboptimal policies for it. In fact, this last analysis constitutes part of a broader initiative of ours, concerning the development of efficient, adaptive policies for the original ONV problem and its variations considered in this work; a first set of results on this problem can be found in [19].

Appendix: Proof of Lemma 1

Let $\psi'_n = \min\{k : S_k > n \cdot c\}$. Then ψ'_n is a stopping time and, from Lemma 2.3 of [16], we have that

$$E[(\sum_{i=1}^{\psi'_n}(X_i - \mu))^r] \quad \leq \quad C(r, E[X^r]) \cdot E[(\psi'_n)^{r/2}] \tag{63}$$

where $C(r, E[X^r])$ is a constant depending only on r and $E[X^r]$. Equation 63 further implies that

$$E[n^{-r/2} \cdot (\sum_{i=1}^{\psi'_n}(X_i - \mu))^r] \quad \leq \quad C(r, E[X^r]) \cdot E[(\frac{\psi'_n}{n})^{r/2}] \tag{64}$$

From Equation 64 and Theorem 2.3 of [16], we get

$$\sup_{n \geq 1} E[n^{-r/2} \cdot (\sum_{i=1}^{\psi'_n}(X_i - \mu))^r] < \infty \tag{65}$$

which implies the uniform integrability of $\{n^{-r/2} \cdot (\sum_{i=1}^{\psi'_n}(X_i - \mu))^r, n \geq 1\}$ [20].

By the definition of the renewal process ψ'_n,

$$n \cdot c = \sum_{i=1}^{\psi'_n} X_i + (\sum_{i=1}^{\psi'_n} X_i - n \cdot c) \tag{66}$$

which further implies that

$$n^{-1/2} \cdot (n \cdot c - \mu \cdot \psi'_n) = n^{-1/2} \cdot \sum_{i=1}^{\psi'_n} (X_i - \mu) + n^{-1/2} \cdot (\sum_{i=1}^{\psi'_n} X_i - n \cdot c) \tag{67}$$

Equation 67 combined with the triangle inequality and the fact that

$$0 \leq \sum_{i=1}^{\psi'_n} X_i - n \cdot c \leq K \tag{68}$$

also imply that

$$|n^{-1/2} \cdot (n \cdot c - \mu \cdot \psi'_n)| \leq |n^{-1/2} \cdot \sum_{i=1}^{\psi'_n} (X_i - \mu)| + n^{-1/2} \cdot K \tag{69}$$

and based on the inequality $(a+b)^r \leq 2^{r-1} \cdot (|a|^r + |b|^r),\ a, b \in R$, we finally get

$$|n^{-1/2}(n \cdot c - \mu \cdot \psi'_n)|^r \leq 2^{r-1} \cdot (|n^{-1/2} \sum_{i=1}^{\psi'_n} (X_i - \mu)|^r + n^{-r/2} \cdot K^r) \tag{70}$$

Hence, the uniform integrability of $\{n^{-r/2} \cdot (\sum_{i=1}^{\psi'_n} (X_i - \mu))^r, n \geq 1\}$ and Equation 70 imply the uniform integrability of $\{n^{-r/2} \cdot (n \cdot c - \mu \cdot \psi'_n)^r, n \geq 1\}$. Since $\psi'_n = \psi_n + 1$ we have that

$$n^{-1/2} \cdot (n \cdot c - \mu \cdot \psi_n) = n^{-1/2} \cdot (n \cdot c - \mu \cdot \psi'_n) + n^{-1/2} \cdot \mu \tag{71}$$

which gives

$$n^{-r/2} \cdot |n \cdot c - \mu \cdot \psi_n|^r \leq 2^{r-1} \cdot (n^{-r/2} \cdot |n \cdot c - \mu \cdot \psi'_n|^r + n^{-r/2} \cdot \mu^r) \tag{72}$$

and implies the uniform integrability of $\{n^{-r/2} \cdot (n \cdot c - \mu \cdot \psi_n)^r,\ n \geq 1\}$.

Acknowledgement

This work was partially supported by NSF grants DMI-MES-0318657 and CMMI-0619978.

References

[1] T. Bountourelis and S. Reveliotis, "Optimal node visitation in acyclic stochastic digraphs," in *Proceedings the 8th Intl Workshop on Discete Event Systems (WODES'06).* IFAC, 2006, pp. 358–365.

[2] D. P. Bertsekas, *Dynamic Programming and Optimal Control (3rd ed.).* Belmont, MA: Athena Scientific, 2005.

[3] S. A. Reveliotis, "Uncertainty management in optimal disassembly planning through learning-based strategies," *IIE Trans.*, vol. 39, pp. 645–658, 2007.

[4] S. A. Reveliotis and T. Bountourelis, "Efficient PAC learning for episodic tasks with acyclic state spaces," *Journal of Discrete Event Systems: Theory and Applications*, vol. 17, pp. 307–327, 2007.

[5] ——, "Optimal flow control in acyclic networks with uncontrollable routings and precedence constraints," School of Industrial & Systems Eng., Georgia Tech (under review in IEEE Trans. on Automatic Control), Tech. Rep., 2008.

[6] S. M. Ross, *Stochastic Processes*. NY: Wiley and Sons, 1996.

[7] C. H. Papadimitriou, "Games against nature," *Journal of Computer and System Sciences*, vol. 31, pp. 288–301, 1985.

[8] J. Niño–Mora, "Stochastic scheduling," in *Encyclopedia of Optimization*, C. A. Floudas and P. M. Pardalos, Eds. Kluwer, 2001, pp. 367–372.

[9] M. Pinedo, *Scheduling: Theory, Algorithms and Systems (2nd ed.)*. Upper Saddle River, NJ: Prentice Hall, 2002.

[10] J. G. Dai, "Stability of fluid and stochastic processing networks,," Center for Mathematical Physics and Stochastics, University of Aarhus, Denmark, Tech. Rep. ISSN 1398-7957, 1999.

[11] H. Chen and D. D. Yao, *Fundamentals of Queueing Networks: Performance, Asymptotics, and Optimization*. NY,NY: Sprigner, 2001.

[12] S. Meyn, *Control Techniques for Complex Networks*. Cambridge, UK: Cambridge University Press, 2008.

[13] D. Bertsimas and D. Gamarnik, "Asymptotically optimal algorithms for job shop scheduling and packet switching," *Journal of Algorithms*, vol. 33, pp. 296–318, 1999.

[14] D. Bertsimas and J. Sethuraman, "From fluid relaxations to practical algorithms for job shop scheduling: The makespan objective," *Mathematical Programming*, vol. 92, pp. 61–102, 2002.

[15] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*. Belmont, MA: Athena Scientific, 1997.

[16] A. Gut, "On the moments and limit distibutions of some first passage times," *The Annals of Probability*, vol. 2, No. 2, pp. 277–308, 1974.

[17] D. P. Bertsekas, *Nonlinear Programming, 2nd ed.* Belmont, MA: Athena Scientific, 1999.

[18] T. Bountourelis and S. A. Reveliotis, "Customized learning algorithms for episodic tasks with acyclic state spaces," School of Industrial & Systems Eng., Georgia Tech, Tech. Rep., 2008.

[19] T. Bountourelis and S. Reveliotis, "Rollout policies for the problem of optimal node visitation in acyclic stochastic digraphs," in *European Control Conference 2007*. IEEE, 2007, pp. 2456–2463.

[20] P. Billingsley, *Convergence of probability measures*. NY: Wiley and Sons, 1968.