AFRL-AFOSR-JP-TR-2018-0064

Negotiating Mission Plans under Risk Bounds

Sylvie Thiebaux
COMMONWEALTH SCIENTIFIC AND INUSTRIAL RESEARCH ORGANISATION
LIMEEONE AVE
CABERRA, 2612
AU

08/28/2018
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
Air Force Office of Scientific Research
Asian Office of Aerospace Research and Development
Unit 45002, APO AP 96338-5002

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
28-08-2018

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
14 May 2015 to 13 May 2018

4. TITLE AND SUBTITLE
Negotiating Mission Plans under Risk Bounds

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA2386-15-1-4015

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Sylvie Thiebaux

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
COMMONWEALTH SCIENTIFIC AND INUSTRIAL RESEARCH ORGANISATION
LIMEEONE AVE
CABERRA, 2612 AU

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AOARD
UNIT 45002
APO AP 96338-5002

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR IOA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-JP-TR-2018-0064

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The research group headed by Professor Sylvie Thiebaux, made significant advances in the development of practical solutions to constrained stochastic shortest path problems (C-SSPs). A specific problem of interest that can be modelled as a C-SSP is that of autonomous systems operating in uncertain environments, facing the problem of optimizing their performance whilst satisfying complex mission constraints with high probability and bounding the risk of plan execution failure. This is currently an active research topic in the operations research, artificial intelligence, robotics, and software verification communities. However, all existing algorithms for C-SSPs require generating and exploring the entire state space of the problem, making them impractical for autonomous systems which have huge state spaces.
Over the course of this research project, the group has published ten top tier publications, two of which received best paper awards at the International Conference on Automated Planning and Scheduling (ICAPS) in 2016 and 2017. A summary of their results search group is highlighted by their development of the first heuristic search algorithms for C-SSPs. These algorithms typically explore a small fraction of the state space, and enable solving much larger problems than was previously possible. To be effective, they must be provided with an admissible heuristic function, that is a lower bound on the expected cost to reach the system goal under the constraints. To achieve this, the research group devised the first domain-independent heuristics that take into account uncertainty, costs, and constraints. Their heuristics have since become the state of the art also for regular (unconstrained) SSPs: even though heuristic search had been used to solve SSPs for over two decades, existing heuristics ignored uncertainty altogether.
Moreover, Professor Thiebauxs group exten

15. SUBJECT TERMS
Automated Planning

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			SINGLETON, BRIANA

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Unclassified	Unclassified	Unclassified	SAR	**PAGES**	**19b. TELEPHONE NUMBER** *(Include area code)* 315-227-7007

Final Report for AOARD Grant FA2386-15-1-4015

Negotiating Mission Plans under Risk Bounds

(Generating Mission Plans Under Risk Bounds*)

Sylvie Thiébaux (PI), **Patrik Haslum** (Co-PI)
Peter Baumgartner, Felipe Trevizan
sylvie.thiebaux@anu.edu.au
Data61 & The Australian National University
Bld 145, ANU, Acton 2601, Australia
T: +61 2 6125 8678

Brian Williams (Co-PI)
Pedro Santana, Tiago Vaquero
williams@csail.mit.edu
Massachussetts Institute of Technology
32 Vassar St., Cambridge, MA 02139, USA
T: +1 617 253-2739

11 August 2018

Abstract

Autonomous systems operating in uncertain environments face the problem of optimizing their performance whilst satisfying complex mission constraints with high probability and bounding the risk of plan execution failure. Such problems can be modelled as constrained stochastic shortest path problems (C-SSPs), which are an active research topic in the operations research, artificial intelligence, robotics, and software verification communities. However, all existing algorithms for C-SSPs require generating and exploring the entire state space of the problem, making them impractical for autonomous systems which have huge state spaces.

This project has made significant advances towards more practical solutions to C-SSPs. We have devised the first heuristic search algorithms for C-SSPs. These algorithms typically explore a small fraction of the state space, and enable solving much larger problems than was previously possible. To be effective, they must be provided with an admissible heuristic function, that is a lower bound on the expected cost to reach the system goal under the constraints. To achieve this, we have devised the first domain-independent heuristics that take into account uncertainty, costs, and constraints. Our heuristics have become the state of the art also for regular (unconstrained) SSPs: even though heuristic search had been used to solve SSPs for over two decades, existing heuristics ignored uncertainty altogether.

Moreover, we have extended these algorithms and heuristics in several directions, including to rich probabilistic temporal logic constraints, to partially observable environments, to hybrid discrete/continuous environments, and to efficiently handle dead-ends. In doing so, we have bridged the gap between several areas of research, across several scientific communities, specifically, heuristic search, classical planning heuristics and planning under uncertainty in Artificial Intelligence, Markov decision processes in Operations Research, synthesis of policies with probabilistic temporal logic objectives in the Formal Verification community, and motion planning in Robotics.

Two of the ten top tier project publications have received best paper awards at the International Conference on Automated Planning and Scheduling (ICAPS) in 2016, and 2017, respectively. Applications of these results are underway in collaboration with partners in the aerospace and automobile industries.

*The original title of the project was *Negotiating Mission Plans under Risk Bounds*. We take the liberty to modify this to reflect the actual emphasis of the work done.

Contents

Chapter 1

Report

Introduction

Motivation

The general problem we consider is automated planning and execution for autonomous systems in uncertain environments for which a probabilistic model is available. These problems are traditionally modelled as goal-oriented (partially observable) Markov decision processes, also known as Stochastic Shortest Paths problems (SSP). A solution to an SSP is a plan that is robust in the face of uncertainty, and which takes the form of a policy mapping states (or state histories) to actions. In almost any realistic environment however, system constraints and mission constraints specified by the user cannot be guaranteed with certainty and no goal can be achieved without taking some risk. The objective of the project is to develop algorithms for generating optimal (or good) policies that bound the risk of undesirable behaviors and outcomes.

Sensitivity to uncertainty and risk of failure is key in enabling autonomous agents to reliably complete missions in real-world environments. Moreover, handling risk is required for automated planning algorithms to be trusted by mission planners, and is a prerequisite to their wider deployment within autonomous systems. This need is more evident in safety-critical application domains such as autonomous planetary or underwater exploration, in which optimizing performance is not enough, and utility must be traded-off against the risk of jeopardizing mission goals. For instance, a Mars rover should optimize the value of the gathered scientific information and information sent to earth, while being subject to safety requirements and constraints such as avoiding collision against obstacles, not traversing into unsafe terrains, maintaining safe operational temperature and battery levels. In these domains, autonomous agents should seek to optimize expected science return while remaining safe by deliberately and proactively keeping the probability of violating one or more constraints within acceptable levels.

In order to model such planning problems, we consider *constrained* SSPs (C-SSPs), that is SSPs with chance constraints: rougly speaking, constraints that are satisfied with (at least) a given probability θ. The simplest chance constraints express that the system should reach a goal state and/or avoid undesirable states with at least a certain probability. Slightly more complex are quality of service constraints expressing that the expected consumption of a certain resource during the plan execution must not exceed a certain threshold. We are interested in these type of constraints, but also in a richer class of chance constraints, described by probabilistic temporal logic formulae. These formulae enable expressing constraints on the probabilities and the temporal properties of paths of the system. An example would be "with probability 0.98, if the temperature exceeds 25 degrees C, then within 10 minutes return to a temperature of 24 degrees and keep it there for at least 30 minutes".

Existing Work

These constrained SSPs have been an active research topic in Operations Research, but also in the Formal Verification and Robotics where they are used to synthesize controllers for probabilistic environments under constraints. In artificial intelligence, C-SSPs have only more recently become a topic of investigation. There is a flurry of efficient AI algorithms that generate optimal policies for ordinary SSPs – for instance algorithms based on heuristic search or Monte Carlo tree search – but the same is not true for C-SSPs. Prior to this project, the state of the art C-SSP algorithm, even with the simpler form of chance constraints above was linear programming, a method that requires enumerating the entire state space of the SSP – something that cannot even fit in memory for anything but toy planning problems. This is further exacerbated for C-SSPs with probabilistic temporal logic constraints. The state of the art algorithms developed by the Formal verification community apply linear programming to the cross-product of the SSP and a (Rabin) automaton representing the extra memory required to decide whether the temporal logic formula holds. This leads to fully enumerating an exponentially larger state space. To cope with complexity, the robotics community has focused on approximate but more efficient solutions to C-SSPs with subclasses of probabilistic temporal logic constraints, including in an on-line planning and execution setting.

Project Goals and Contributions

Our overall aim is to enable the practical generation of optimal policies in the presence of a rich set of chance constraints, and the use of these algorithms when planning risk-bounded missions for autonomous vehicles, planetary rovers, and energy systems.

Our main technical goal is to develop efficient algorithms for C-SSPs with probability of failure, quality of service, and probabilistic temporal logic constraints, which only enumerate a small fraction of the state space. To achieve this, we leverage heuristic search approaches – arguably the most successful approaches for classical planning and ordinary SSPs – and extend them to the more complex constrained SSP case.

In doing so, we advance the state of the art and bridge the gap between AI and adjacent communities, such as Operations Research, Robotics, and Formal Verification.

The main outcomes of this project are:

- i-dual, and i^2-dual, the first heuristic search algorithm for C-SSPs. i^2-dual is the current state of the art for C-SSP.
- The first heuristics for SSPs and C-SSPs taking probability and costs into account (as well as constraints). They are the current state of the art for SSPs and C-SSPs.
- The first heuristic search algorithm for C-SSP with (multi-objective) probabilistic temporal logic constraints. It solves problems out of reach of the well-established PRISM model-checker.
- The first correct and complete algorithm for MDPs with constraints in the very expressive probabilistic temporal logic PCL*.
- RAO* the first heuristic search algorithm for partially observable C-SSPs. RAO* is capable of solving problems of practical significance.
- An extension of RAO* to hybrid systems and applications to space and autonomous vehicles.

Report Organization

Since the project led to a substantial number of publications which adequately describe its outcomes and technical contributions, we provide, in the next section, a rather informal overview of these outcomes, and their novelty and significance. We then discuss future work motivated or enabled by the project results. Finally, in the appendix, we directly include, the most relevant publications that describe our results in detail. References to the literature are provided in each of the publications, rather than in the overview part of this report.

Overview of Main Results

Outcome 1: i-dual – Heuristic Search for Constrained SSPs

The first outcome of the project is an efficient algorithm to solve C-SSPs with chance constraints bounding the risk (probability) of reaching a desirable/undesirable state, and/or the risk of not meeting a quality of service indicator such as an expected cost/resource/duration bound.

This algorithm, i-dual (Incremental Dual), is inspired by the efficient heuristic search algorithms the AI community has developed for ordinary SSPs, such as LAO* or (L)RTDP. These search towards a policy minimizing the expected cost to reach the goal, guided by heuristic functions (lower bound on the expected cost) provided by the problem modeler or automatically extracted from the planning domain description. Heuristic guidance enables the explicit generation and exploration of a fraction of the huge state space for the problem. For that reason, heuristic search algorithms can solve much larger problems than linear programming or dynamic programming methods for SSPs.

However, adapting these heuristic search algorithms to C-SSPs is not an easy task. One complication is that these algorithms search in the space of deterministic policies. This is sufficient for SSPs, because optimal SSP policies are deterministic, but not for C-SSPs as optimal policies for C-SSPs can be stochastic. We therefore need to search in the infinite space of stochastic policies. Another complication is that the policy must satisfy the chance constraints, which implies representing them and checking for their satisfaction efficiently.

To resolve these issues i-dual combine the strengths of heuristic search and linear programming in the space of occupation measures (these are the variables of the LP and they represent the expected number of times an action will be taken in a given state). This is the dual space space of that considered by existing heuristic search methods for ordinary SSPs (these methods operate in the space of value functions). As other heuristic search algorithms, i-dual progressively expand the reach of the best policy found so far and heuristically evaluates the fringe of the policy. However, it updates this best policy using linear programming in the occupation measure space. It uses heuristics for the cost to be optimized, and also for the costs/probabilities to be bounded by the chance constraints. The latter enable early pruning of policies violating the chance constraints. Interestingly, if the former heuristic is not admissible, i-dual can obtain a sub-optimal (but valid) policy quickly – an option that is not offered at all by linear programming.

We implemented a domain-independent planner based on i-dual, taking as input a variant of the standard Probabilistic Planning Domain Definition (PPDDL) language. We found that it is able to solve benchmark problems up to two orders of magnitude faster than linear programming and to easily solve problems which are much too large to be formulated as an LP (e.g. problems that would require many billions of LP variables).

I-dual is novel and significant because:

- it extends the reach of heuristic search methods:
  - as it is the first polynomial-time heuristic search algorithm solving C-SSPs optimally;
- it is the first optimal algorithm for C-SSPs that has a chance of being practical:
  - as it is the first method for C-SSPs that does not enumerate the state space;
- it opens the field of planning under uncertainty to a different class of approaches:
  - as it searches through a different state space than existing SSP heuristic search methods, and bridges the gap between AI and OR approaches to SSPs.

Outputs for i-dual:

- Our work on i-dual received the Best Paper Award the 2016 International Conference on Automated Planning and Scheduling (ICAPS-16), the premier conference in the field.
- A short paper on i-dual was invited for publication and presentation at the 2017 International Joint Conference on Artificial Intelligence (IJCAI) Sister Conference Best Paper Track.
- i-dual was re-implemented by Joërg Hoffmann and students and incorporated into Fast Downward, one of the most used planning systems.

More information: see Appendix A and papers [10] and [6].

Outcome 2: Occupation Measure Heuristics for (Constrained) SSPs

Clearly i-dual's effectiveness in guiding the search depends on the informativeness of the heuristic function. So a major part of our efforts has been spent on the automatic design of heuristics for C-SSPs.

The classical planning community has very-well developed theories and algorithms to automatically generate good heuristics from the description of a planning problem to be solved. In particular, admissible heuristics (lower bounds) are obtained by optimally solving a relaxation of the planning problem in low-order polynomial time and many such relaxations have been studied. However, current heuristics assume that the problem is deterministic. For SSPs, the state-of-the art heuristics first determinize the problem (relaxing it by splitting every probabilistic action with n outcomes into n deterministic actions), and apply a known deterministic heuristic to this deterministic problem. The major drawback of this approach is that the resulting heuristics completely ignore probabilities as well as the tradeoff between probabilities and costs. Doing better has been an open problem for the past 25 years.

To remedy this situation, we focused on the occupation measure space explored by i-dual. Occupation measures represent the expected number of times a given action will be executed in a given state of a policy. By relaxing the linear program that optimizes them, we derive occupation measure heuristics, the first admissible heuristics for SSPs and C-SSPs that take probabilities into account. We show that these heuristics are the probabilistic counterpart of the "operator-counting" heuristics recently introduced in classical planning, and that they can be obtained by incorporating probabilistic information into the latter, establishing a bridge with research in classical planning. In our experiments, we found that occupation heuristics can substantially improve the number of expanded nodes, the run-time and number of problems solved in comparison with existing deterministic heuristics. This holds regardless of the search algorithm used, and for both C-SSPs (using i-dual) and regular SSPs (using algorithms like LAO*, LRTDP and i-dual).

Moreover, one of the most intriguing advantages of occupation measure heuristics is that they can be computed at once for multiple states, using the same set of linear constraints. Thus, their formulation can directly be incorporated into the LP solved by i-dual to update the policy at each iteration. This leads to i^2-dual, a brand new type of heuristic search method for C-SSPs where the heuristic computation is lazy, reusable across multiple parts of the search space, and works in unison with the policy update. For the C-SSP benchmarks we experimented with, we found that i^2-dual improves over i-dual regardless of the heuristic used by the latter. It solved more problems than the latter and improved the run-time by over an order of magnitude on some problems.

Occupation measure heuristics are novel and significant because:

- they are the first heuristics that take both probabilities and costs into account
- they solve a long-standing open problem in probabilistic planning
- they allow larger (C-)SSPs to be solved optimally (and smaller ones to be solved faster)
- they naturally extend the recent deterministic "operator-counting" heuristics to probabilities
- they can be formulated to compute the values of multiple state at once
- their computation can be fully integrated with policy update, leading to a new type of algorithm for probabilistic planning and to the new state of the art for C-SSPs: i^2 -dual

Outputs for occupation measure heuristics:

- In virtue of its significance and novelty, our work on occupation measure heuristics received the Best Paper Award the 2017 International Conference on Automated Planning and Scheduling (ICAPS-17).
- A short paper on these heuristics was invited for publication and presentation at the 2018 International Joint Conference on Artificial Intelligence (IJCAI) Sister Conference Best Paper Track.

More information: see Appendix B and papers [8] and [4].

Outcome 3: Solving SSPs with Probabilistic Temporal Logic Constraints

One the key objectives of this project was to handle more expressive mission constraints expressed in probabilistic temporal logic, to generate policies providing probabilistic guarantees about the properties of the sequence of states followed by the autonomous system. For instance, a policy for a search and rescue UAV mission might need to minimize the expected time to get survivors to safety, whilst avoiding dangerous areas at all times, and repeatedly circling the affected locations for a given period of time to correctly determine the presence of survivors with high probability. Optimal policies for such problems are not only stochastic but also history dependent, so as to capture the temporal nature of the constraints.

Policy synthesis algorithms for Markov decision processes with probabilistic temporal logics such as probabilistic LTL, probabilistic CTL and their parent probabilistic CTL*, stem from the area of probabilistic verification. They struggle with the large state spaces and constraint types found in automated planning. This is partly due to properties of the problem (undecidable for PCTL*, 2-EXPTIME complete for PLTL), but also to their inherent efficiency. In particular such algorithms compute rabin automata of double exponential size in the size of the formulae, synchronize them with the entire state space of the system, and run vanilla linear programming on the resulting cross product to obtain a policy. Given that planning problems have gigantic state spaces, this approach is obviously inapplicable to anything but tiny problems.

Our approach differs from the state of the art in two crucial ways. Firstly it operates entirely on-the-fly, bypassing the expensive construction of Rabin automata for the formulas and their prohibitive prior synchronization with the full state space of the SSP. We are able to represent the policy history using modes computed from non-deterministic Büchi automata which we determinize on-the-fly or from formula obtained by progression of the original temporal logic formula – both are exponentially smaller than Rabin automata. Secondly, our approach extends our heuristic search algorithm i^2-dual, and our occupation measure heuristics to enable pruning the regions made infeasible by the probabilistic temporal logic constraints. Our experiments show that PLTL-dual, the resulting algorithm, is capable of solving planning problems that are clearly out of reach of the state-of-the-art Prism probabilistic model-checker.

To achieve a compromise between efficiency and expressivness, our work has focused on handling SSPs with multi-objective probabilistic LTL (MO-PLTL) constraints, i.e. conjunctions of probabilistic LTL formulas. Such problems can be compiled into C-SSPs with an extended state space which captures the history relevant to determining the satisfaction of the constraints. An optimal solution takes the form of a finite-memory stochastic policy whose execution: (a) satisfies the MO-PLTL constraints; (b) reaches a goal state with probability 1; and (c) has minimal expected cost, subject to (a,b). We have proven that PLTL-dual correct and optimal for such constraints, showed that it has reduced worst-case complexity due to the presence of a goal (SSP vs MDP), and demonstrated encouraging scalability results.

PLTL-dual is novel and significant because:

- it is the first heuristic search algorithm for SSPs with probabilistic temporal logic
- its worst case complexity improves over the state of the art (single vs double exponential)
- its practical performance on planning problems exceeds that of the well developed model-checker PRISM

Outputs for PLTL-dual:

- Our work on heuristic search for SSPs with multi-objective probabilistic LTL constraints has been accepted for presentation at the 2018 International Conference on Principles of Knowledge Representation and Reasoning (KR-18).

More information: see Appendix C and paper [3].

Outcome 4: RAO*: Risk-Aware Planning with Partially Observability

Many real-world domains are partially observable either because sensors do not exist that observe a variable that is relevant to making optimal decisions, or because these sensors are not entirely reliable. To handle these domains, RAO* (Risk-bounded AO*), which predates[1] i-dual, is focused on *partially observable* C-SSPs (C-POMDPs)[2], which are much more difficult than SSPs. In particular, partial observability implies searching in the space of belief states (probability distributions on states), rather than in the mere space of system states.

RAO* is an extension of the well-known AO* heuristic search algorithm. It is the first heuristic search algorithm capable of generating optimal deterministic policies for solving finite-horizon C-POMDPs with chance constraints bounding the risk of failure (requiring probability of goal satisfaction to be above a certain threshold). Our work on RAO* showed that existing approaches combining chance constraints and POMDPs are incorrect unless the violation of a chance constraint is observable and causes the plan execution to halt. These hypotheses are not satisfied by many real-world problems and RAO* does not require them. One problem that motivated RAO* is power supply restoration following a fault on a power distribution system. When restoring power, one wants to avoid, with high probability, to re-supply a fault as this could be dangerous and even potentially cause the loss of human lives – however, there is always a risk that this will happen as the fault location is uncertain due to sensor uncertainty. Yet, this should not cause the policy execution to halt.

On the one hand, RAO* may seem more limited than i-dual in that it handles a smaller class of chance constraints, generates deterministic policies even though the optimal policy may be stochastic, and requires a finite planning horizon. On the other hand, the two last restrictions may be desirable from a practical point of view, and furthermore RAO* does handle partial observability which we have not investigated with i-dual. Our implementation of RAO* produced astonishingly good results on difficult problems such as power supply restoration. These encouraging results led us to start demonstrating its use on practical problems.

Specifically, MIT used RAO* to generate conditional plans for the control of a simulated Mars Rover, in joint collaboration with JPL and Caltech. The purpose of this demonstration is to illustrate improved mission robustness and autonomy, through the execution of conditional plans that guarantee bounded risk of failure. MIT also used RAO* to demonstrate the automated synthesis of recovery procedures for manufacturing robotics, in collaboration with Mitsubishi, and for a driver's assistant, in collaboration with Toyota. With respect to manufacturing, a major cost in deploying manufacturing robots is the development of procedures that account for the most likely system failures. These procedures are automatically generated by RAO*. In the area of automobile safety, a source of accidents is driver distraction, for example when congestion causes a driver to be late for a meeting. We employ RAO* to generate contingency procedures for likely outcomes, and to help the driver to anticipate these outcomes and contingencies, before a major failure occurs.

RAO* is novel and significant because:

- it is the first heuristic search algorithm for optimally solving C-POMDPs,
- it correctly handles risk and chance constraints in the presence of partial observability,
- it can be efficiently implemented and used in a range of practical settings.

Outputs for RAO*:

- Our work on RAO* was published and presented at the 2016 AAAI Conference on Artificial Intelligence (AAAI-16).

More information: see Appendix D and paper [11].

[1] 1 Research on RAO* started as part of AFOSR grant FA9550-12-1-0348 (PI Brian Williams).

[2] Even though these problems have a goal state and should therefore be called POSSPs, we use the well-known term POMDP because POSSP isn't used in the literature.

Outcome 5: Risk-Aware Planning in Hybrid Domains and Applications

Among the numerous requirements that a fully autonomous system must meet in real problems, three of them are key enabling capabilities for exploration missions, whether in space or in the ocean: 1) the ability to act under uncertainty and adapt to its environment; 2) the ability to optimize performance while offering hard bounds on the risk of mission failure or mission constraints not being met; and 3) the ability to handle complex hybrid (discrete and continuous) planning problems in a provably correct and scalable fashion. In the first part of the project, we we focused on addressing the first two capabilities with RAO* and i-dual. During the second part we leveraged the work on RAO* to address chance constraints planning in hybrid domains, in which activity planning with C-POMDP models is coordinated with risk-bounded scheduling and risk-bounded path planning.

Inspired by several Mars Rover scenarios from MIT's collaboration with JPL, we developed a hybrid conditional planner, CLARK, to support time-critical missions with temporal uncertainty on the vehicle's activity duration, limited communication time windows and uncertainty on navigating the environment with obstacles. CLARK uses RAO* as its core chance constrained planning algorithm to reason on a discrete C-POMDP model. Each potential actions in the C-POMDP requires the planner to reason on the continuous world, by checking the temporal consistency of the plan and also synthesizing the vehicle trajectory according to the risk bounds. In order to achieve this, CLARK manages the interface between RAO* and two subsolvers. The first solver, PARIS handles the risk-bounded temporal reasoning and is the fastest algorithm for risk-minimal and chance-constrained strong scheduling of SNTUs (Simple Temporal Networks with Uncertainty) and PSTNs (Probabilistic STNs). The second solver, pSulu was previously developed by the MIT team to handle risk-bounded motion planning. CLARK is the first planner capable of fully handling a coordination of chance constrained activity planning, probabilistic scheduling and path planning with uncertain dynamics.

The CLARK planner has been demonstrated in a set of simulated Mars rover scenarios provided by JPL collaborators, in which the autonomous vehicle has to navigate a map with obstacles containing five locations of interest and perform a series of data collection activities. The rover is given a set of science requests, including 1) taking pictures from pre-specified locations; and 2) collecting rock samples from potential collection sites at locations believed to contain rocks with high probability. Upon completion of its data collection (both taking pictures and collecting rock samples), the rover is required to drive to a location from which its science data can be transmitted to an orbiting satellite. The orbiter is visible from these locations within limited windows of time. When the data from each request is transmitted, the request is deemed completed. In this problem, our planner has to constrain the chance of missing the time window for transmitting the requested data and the chance of collision while moving between target locations.

We also pursued a complementary approach in which we directly extended RAO* to handle hybrid domains in an on-line fashion, resulting in i-RAO* (incremental RAO*), a risk-aware heuristic forward search algorithm producing optimal deterministic, finite-horizon policies for constrained POMDPs with continuous-discrete state and observation. The search explores a risk-aware AND-OR tree in continuous and discrete belief space that describes the optimal polices conditioned on discrete observations. To update the belief states, the distribution of continuous states is obtained in advance based on action and unknown future continuous observation using Linear Quadratic Gaussian Motion Planning (LQG-MP), whilst the distribution of discrete sates is updated using the state transition model and observation. We also extended this heuristic search framework to durative actions, where each action is a maneuver over a specific time interval.

We have demonstrated the usefulness of i-RAO* on the Mars rover scenario as well as in planning maneuvers for autonomous vehicles. In the latter problem, off-line computations include generating a library of probabilistic maneuvers for the controllable agent and planning an initial motion policy to execute. During runtime, the conditional planner can quickly look up maneuver sequences to ensure risk bounds as the world around our agent evolves.

Outputs for PARIS, CLARK and i-RAO*:

- Our work on PARIS was published and presented at the 2016 International Conference on Automated Planning and Scheduling (ICAPS-16).
- Our work on CLARK and its application to the planetary rover scenarios was presented at the

2016 Space and Astronautics Forum and Exposition of the American Institute for Aeronautics and Astronautics (AIAA SPACE).

- Our work on i-RAO* and its autonomous vehicles application has been accepted for publication at the 2018 IEEE Conference on Decision and Control (CDC-18).

More information: see Appendix E and papers [9] [12] and [2].

Other Notable Results

Outcome 6: Improving i-dual's Efficiency with Column Generation

As explained above, one of the most important factors governing the efficiency of i-dual is the informativeness of the heuristic. The second way of improving the efficiency of i-dual is to improve some of the choices made by the algorithm. To solve C-SSPs, i-dual applies linear programming in the occupation measure space to a small subset of the reachable state space rooted at the initial state. At each iteration, it grows the subset of states considered by adding all successors of all (non-goal) fringe states reachable from the initial state under the best policy found by the linear program. This goes on until all of these reachable fringe states are goal states, at which point i-dual terminates. One idea to improve the efficiency of i-dual is to only generate and add only some of the successors of the fringe states, selecting those likely to improve the policy's expected cost.

In the language of operations research, i-dual can be viewed as a kind of column generation algorithm. That is, an algorithm that solves increasingly larger subproblems of the original problem, adding new variables at each iteration (in our case the new variables are occupation measures for the new states generated). The operations research community has well-established ways of selecting these new variables in order to produce the best possible improvement to the objective function, by looking at the "reduced cost" of the variables.

We have therefore investigated whether we could improve the efficiency of i-dual by extending these reduced-cost-based strategies to our setting, to choose the best successors of the fringe states to add to the search space, that is, which occupation measure variables to add to our LP. Our case is more complex than the traditional column generation setting as a) the cost of the fringe states (hence the objective function of the LP and some of the constraints) is heuristically estimated, and b) not all constraints are present in the subproblems we are solving. In particular, flow constraints for the occupation measures associated with the states that haven't yet been generated are absent. Nevertheless, in our experiments, a reduced-cost based selection of the new variables to add resulted in run-time improvements of x2-x5 in some cases.

We have been invited to submit an extension of our ICAPS-16 paper on i-dual by the Artificial Intelligence journal, arguably the best journal in AI. Our submission incorporates the column-generation results.

Outcome 7: MCMP: Efficiently Handling SSPs with Dead-Ends

Many (C-)SSPs have dead-ends, that is, states from which the problem goal isn't reachable. In those circumstances, one would ideally want to maximize the probability of success and minimize the expected cost to reach the goal. The ad-hoc way of handling these conflicting objectives is to assign a (large) finite fixed cost/penalty to each dead-end state. This results in policies that give up trying to achieve the goal as soon as the expected cost accumulated so far exceeds the dead-end penalty. This solution is unsatisfactory in that it is difficult to determine a penalty that leads to the desired behavior.

An alternative approach is to optimize a dual criterion taking into account both the goal satisfaction probability and the expected cost. Existing models that camp include S3P and iSSPUDE: they minimize the cost of the execution trajectories leading to the goal, ignoring the others. They have two drawbacks: they do not account for the cost of trajectories that result in dead-ends, and algorithms for solving them are inefficient in comparison to algorithms for SSPs without dead-ends. Using C-SSPs and i-dual, we have proposed a new model which doesn't have these drawbacks: MCMP (Min-Cost given Max-Prob). This model leads to the minimum expected cost policy among those with maximum success probability, and accurately accounts for the cost and risk of reaching dead-ends. MCMP is solved in two steps: first find out the maximum probability of success, which can be done efficiently using a slight variant of i-dual; second find the policy with minimum expected cost, under the constraint that the probability of success is what was determined in the first step, which is a C-SSP solvable efficiently with i-dual. C-SSPs with dead-ends can be solved similarly, by adding the chance constraints to each step.

Our experiments show two order of magnitude speed up over the best existing algorithm for S3P. This work was published at the 33rd Conference on Uncertainty in Artificial Intelligence (UAI-17). See Appendix F and paper [7] for details.

Outcome 8: C-SSPs with PCTL* Constraints

In handling C-SSP with a wide range of constraints, our methodology was to start with simple constraint forms, such as those bounding the risk of failure or the quality of service and then move on to the more complex constraints in probabilistic temporal logic. There is a tradeoff between the expressivness of the probabilistic temporal logic considered and the complexity of synthesizing policies satisfying constraints.

In order to gain, in the first instance, the deepest possible insight into handling probabilistic temporal logic constraints, we decided to start by considering the very general PCTL* logic, the probabilistic counterpart of the temporal logic CTL*. No algorithm for policy synthesis in this very natural and expressive framework has been developed so far – previous work has been limited to fragments such as PCTL or PLTL. It is known that optimal policies for PCTL* are both history-dependent and stochastic, and that the synthesis problem in its full generality is undecidable. To ensure decidability, we restricted ourselves to policies whose execution history memory is finitely bounded a priori. Yet, policy synthesis is still an open problem, even with this restriction.

We solved this open problem, producing the first algorithm for bounded-memory policy synthesis for MDPs[3] with PCTL* constraints. Our algorithm is described in the form of a tableau calculus that, given an MDP and a PCTL* specification, derives in a non-deterministic way a system of (possibly nonlinear) equations. The solutions of this system, if any, describe the desired (stochastic, finite-memory) policies. We established the correctness, completeness and termination of our method.

One advantage of this approach over existing ones is that it does not require expanding the full state space upfront, nor its cross-product with an automaton representing the formula. Therefore, even though this algorithm for the full PCTL* has to be inefficient in virtue of the complexity of the problem, it paved the way for devising simpler methods for subsets of PCTL*, such as MO-PLTL (see Outcome 3), which could be directly handled in conjunction with heuristic search and exploit i-dual's benefits.

Our policy-synthesis method for PCTL* is significant because: a) it is the first algorithm for this problem, solving an open problem, b) its nature is substantially different from previous methods, and c) it doesn't require expanding the huge state space.

This work was published and presented at the 26th International Conference on Automated Reasoning with Analytic Tableaux and Related Methods (Tableaux-17). See Appendix F and paper [5] for details.

[3]There are no goal states, hence the use of the MDP terminology.

Discussion

This project opened many directions for future work, some of which we explain here.

Extending the Scope of i-dual

This project has focused on constrained SSPs, that is, a constrained Markov decision process where a goal state must be reached with minimum expected action cost, within a finite amount of time. These are also called indefinite horizon MDPs. However, in the standard MDP model, there is no goal but rather the objective of maximizing the expected cumulative reward, over a finite or infinite horizon. We believe that i-dual and the occupation measure heuristics can be extended to these cases. Moreover, they can also be extended to handle different uncertainty models where the probability distribution is only partially known, including imprecise, set valued and bounded probability models. Such extensions and the application of i-dual across the range of stochastic sequential decision models are high priority on our research agenda.

Monte Carlo Tree Search for Constrained MDPs

Monte Carlo tree search is becoming the method of choice for on-line planning under uncertainty, using finite-horizon MDP models. Extending MCTS to constrained MDPs is a worthwhile short term goal. Providing guarantees of convergence to the optimal however is not entirely trivial, nor is guiding MTCS using heuristics. Patrik Haslum and Sylvie Thiébaux have recently been awarded an Australian Research Council grant which will investigate this as part of its agenda.

Generating Optimal Deterministic Policies with i-dual

Another short-term goal is the generation of deterministic policies with i-dual. Decision-makers are much more likely to trust such policies than stochastic ones, and for right or wrong stochastic policies might not be allowed in safety-critical applications. Surprisingly perhaps, since stochastic policies are superior in terms of expected cost, generating deterministic policies is a harder problem: NP-complete vs polynomial in the size of the (huge) state space. It is possible to extend i-dual to generate such policies by using a mixed-integer program rather than an linear program, but the practicality of this approach needs to be investigated in detail.

On-line Planning with Guarantees

Our work on i-dual with column-generation offers a principled way of determining which parts of a partial policy are more fruitful to extend to improve the quality of the policy. Combined with upper-bounds on expected costs (rather than the lower bounds given by the heuristics), this forms the basis for developing continual on-line planning whilst guaranteeing to escape dead-ends. We're currently working on finding such upper-bounds.

Efficiently Allowing Infinite Behaviors

On the one hand, we have shown how to efficiently handle C-SSP with multi-objective probabilistic LTL constraints. These problems have a goal and do not admit infinite behaviors as the goal must be reached within finite time after which the process ends. For instance, they can't capture that the agent must alternate indefinitely between two rooms. On the other hand, we have also devised a correct and complete algorithm for C-MDP with the full CTL* probabilistic logic constraints. These problems do admit infinite behaviors but their complexity is prohibitive. The formal Verification community typically studies C-MDPs with the less expressive probabilistic (multi-objective) LTL, which still offers the ability to generate controllers for infinite behaviors. We would like to extend i-dual and our heuristics to this case. This requires dealing with bottom end components of the MDP which is where such infinite behaviors take place.

Ethical Decision-Making

One of the big issues raised recently regarding AI is the necessity to make decisions that are ethical rather than purely rational in an economic sense. Having considered probabilistic temporal logic constraints, we are in a good position to handle constraints expressible in other logics, including logics that may be suitable to describe the laws of ethics. We can then solve problems that involve optimizing the agent behavior subject to these laws.

Project Outputs and Interactions

Awards

Papers [8] and [10] below obtained the best paper award at the International Conference on Automated Planning and Scheduling (ICAPS), two years in a row in 2016 and 2017.

Publications

a. Papers submitted to peer-reviewed journals:

[1] Felipe Trevizan, Sylvie Thiébaux, Pedro Santana, Brian Williams. **Column Generation and Heuristic Search for Constrained Stochastic Shortest Path Problems.** Submitted to *Artificial Intelligence*, Elsevier.

b. Papers published in peer-reviewed conference proceedings:

[2] Xin Huang, Ashkan Jasour, Matthew Deyo, Andreas Hoffmann, Brian Williams. **Hybrid Risk-Aware Conditional Planning with Applications in Autonomous Vehicles**. *Proceedings of the 57th IEEE Conference on Decision and Control (CDC-18)*, Miami (FL), USA, December 2018. CORE A.

[3] Peter Baumgartner, Sylvie Thiébaux, Felipe Trevizan. **Heuristic Search Planning with Multi-Objective Probabilistic LTL Constraints**. *Proceedings 16th International Conference on Principles of Knowledge Representation and Reasoning (KR-18)*, Tampe, AZ (USA), October 2018. CORE A*.

[4] Felipe Trevizan, Sylvie Thiébaux, Patrik Haslum. **Operator Counting Heuristics for Probabilistic Planning.** *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI-18)*, Sisters Conference Best Paper Track, Stockholm (Sweden), July 2018. Invited submission of a short version of our ICAPS-17 best paper.

[5] Peter Baumgartner, Sylvie Thiébaux, Felipe Trevizan. **Tableaux for Policy Synthesis for MDPs with PCTL* Constraints**. *Proceedings of the 26th International Conference on Automated Reasoning with Analytic Tableaux and Related Methods (Tableaux-17)*, Brasilia (Brazil), September 2017. CORE A.

[6] Felipe Trevizan, Sylvie Thiébaux, Pedro Santana, Brian Williams. **I-dual: Solving Constrained SSPs by Heuristic Search in the Dual Space**. *Proceedings of the 26 th Joint International Conference on Artificial Intelligence (IJCAI-17)*, Sisters Conference Best Paper Track, Melbourne (USA), August 2017. CORE A*. Invited submission of a short version of our ICAPS-16 best paper.

[7] Felipe Trevizan, Florent Teichteil-Königsbuch, and Sylvie Thiébaux. **Efficient Solutions for Stochastic Shortest Path Problems with Dead Ends**. *Proceedings of the 33rd International Conference on Uncertainty in Artificial Intelligence (UAI-17)*, Sydney (Australia), August 2017. CORE A*.

[8] Felipe Trevizan, Sylvie Thiébaux, Patrik Haslum. **Occupation Measure Heuristics for Probabilistic Planning**. *Proceedings of the 27th International Conference on Automated Planning and Scheduling (ICAPS-17)*, AAAI Press, Pittsburgh (USA), June 2017. CORE A*. **Best paper award.**

[9] Pedro Santana, Tiago Vaquero, Claudio Toledo, Cheng Fang, Brian Williams. **PARIS: a Polynomial-Time, Risk-Sensitive Scheduling Algorithm for Probabilistic**

Simple Temporal Networks with Uncertainty. *Proceedings of the 26th International Conference on Automated Planning and Scheduling (ICAPS-16)*, AAAI Press, London (UK), June 2016. CORE A*.

[10] Felipe Trevizan, Sylvie Thiébaux, Pedro Santana and Brian Williams. **Heuristic Search in Dual Space for Constrained Stochastic Shortest Paths Problems**. *Proceedings of the 26th International Conference on Automated Planning and Scheduling (ICAPS-16)*, AAAI Press, London (UK), June 2016. CORE A*. **Best Paper Award.**

[11] Pedro Santana, Sylvie Thiébaux and Brian Williams. **RAO*: an Algorithm for Chance-Constrained POMDP's**. *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI-16)*, AAAI Press, Phoenix, AZ (USA), February 2016. CORE A*.

c. Papers published in non-peer-reviewed journals and conference proceedings:

[12] Pedro Santana, Tiago Vaquero, Eric Timmons, Brian Williams, Catharine McGhan, Richard Murray, Claudio Toledo. **Risk-aware Planning in Hybrid Domains: An Application to Autonomous Planetary Rovers.** *Proceedings of the AIAA Space and Astronautics Forum and Exposition (AIAA SPACE*], Long Beach, CA (USA), September 2016. http://dx.doi.org/10.2514/6.2016-5537

d. Conference presentations without papers:

[13] Invited presentation by Felipe Trevizan. **Occupation Measures: How OR Can Help Planning under Uncertainty**. *ICAPS-18 Workshop on Planning, Search and Optimization*. Delft (Netherlands), June 2018.

[14] Presentation by Pedro Santana, Tiago Vaquero, Eric Timmons, Brian Williams, Catharine McGhan, Richard Murray, Claudio Toledo, C. **Risk-aware Planning in Hybrid Domains**. *System Demonstration track at the 26th International Conference on Automated Planning and Scheduling (ICAPS-16)*, London (UK), June 2016.

[15] Invited presentation by Felipe Trevizan. **I-dual**. *Australian Society for Operations Research (ASOR) Recent Advances Conference*, Melbourne (Australia), March 2016.

[16] Invited presentation by Brian Williams. **Risk-bounded Programming in Hybrid Domains**. *AAAI-16 Workshop on Planning in Hybrid Domains*. Phoenix, AZ (USA), February 2016.

[17] Invited presentation by Sylvie Thiébaux. **Heuristic Search in Dual Space for Constrained Stochastic Shortest Paths Problems**. *INFORMS Annual Meeting*. Philadelphia, PA (USA), November 2015.

e. Other publications:

[18] Pedro Santana. **Dynamic Execution of Temporal Plans with Sensing Actions and Bounded Risk.** PhD Thesis, Dept. of Aeronautics and Astronautics, Massachusetts Institute of Technology (MIT), September 2016

Interactions and Other Significant Collaborations

- The project is a **collaboration between Data61/ANU and MIT**, with Brian Williams as co-PI. Administratively, MIT is a subcontractor, as AOARD could not dispatch funds to two institutions.
- Sylvie Thiébaux hosted a **one-day visit by Brian Lutz (AOARD)** to ANU in May 2016. Brian gave a presentation to ANU staff about AOARD's programs and future plans in partnership with Australian Defense. He also attended presentations and demonstrations from 7 different ANU research groups, including two presentations concerning current AOARD/AFOSR projects led by ANU/Data61 staff. This contributed to a new project being funded (Sergiy Bogomolov). Sylvie also hosted a **1/2 day visit by Scott Robertson, Alan Lin, and Peter Friedland** in April 2018. They attended presentations from 6 ANU/Data61 research groups, including two presentations concerning current AOARD/AFOSR projects led by ANU/Data61 staff.

- Sylvie Thiébaux and colleagues have started a collaboration with **Airbus Central Research and Technology**, France (Florent Teichteil-Königsbuch) about applying C-SSPs and planning under uncertainty to aerospace applications. Prior to this, we jointly published the UAI-17 paper [7] where we show how to correctly handle dead-ends in SSPs and C-SSPs and use i-dual to improve the state of the art for problems with dead-ends.
- Sylvie Thiébaux started discussions with the **Australian Defence Science Technology Group** (DSTG) regarding potential participation in the Trusted Autonomous Systems Cooperative Research Centre to be led by Australian Defense. This has led to several visits to and from the DSTG's Machine Cognition Group (Martin Oxenham) and preliminary discussions on collaborative projects on goal-based autonomy.
- Brian Williams has continued his **collaboration with JPL and Caltech** in the framework of which RAO*, in coordination with a risk-bound path planner and risk-bounded scheduler, has been demonstrated to plan science missions for a simulated Mars rover and, within one particular rover behavior, to generate control policies for the rover's drilling process.
- Brian Williams has started a **collaboration with Toyota** in which RAO* has been applied to generate risk-bounded sequences of maneuvers for simulated autonomous cars.
- Efforts will be made in the coming year to reach out to USAF researchers regarding the outcomes of this project and potential collaborations.

Appendix A

i-dual: Heuristic Search for Stochastic Shortest Path Problems

Paper:

Felipe Trevizan, Sylvie Thiébaux, Pedro Santana and Brian Williams. **Heuristic Search in Dual Space for Constrained Stochastic Shortest Paths Problems**. *Proceedings of the 26th International Conference on Automated Planning and Scheduling (ICAPS-16)*, AAAI Press, London (UK), June 2016. CORE A*. **Best Paper Award.**

Appendix B

Occupation Measure Heuristics

Paper:

Felipe Trevizan, Sylvie Thiébaux, Patrik Haslum. **Occupation Measure Heuristics for Probabilistic Planning**. *Proceedings of the 27th International Conference on Automated Planning and Scheduling (ICAPS-17)*, AAAI Press, Pittsburgh (USA), June 2017. CORE A*. **Best paper award.**

Appendix C

Heuristic Search for SSPs with Probabilistic Temporal Logic Constraints

Paper:

Peter Baumgartner, Sylvie Thiébaux, Felipe Trevizan. **Heuristic Search Planning with Multi-Objective Probabilistic LTL Constraints**. *Proceedings 16th International Conference on Principles of Knowledge Representation and Reasoning (KR-18)*, Tampe, AZ (USA), October 2018. CORE A*.

Appendix D

RAO*: Risk-Bounded Planning in Partially Observable Domains

Paper:

Pedro Santana, Sylvie Thiébaux and Brian Williams. **RAO*: an Algorithm for Chance-Constrained POMDP's**. *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI-16)*, AAAI Press, Phoenix, AZ (USA), February 2016. CORE A*.

Appendix E

Risk-Bounded Planning in Hybrid Domains

Papers:

Pedro Santana, Tiago Vaquero, Claudio Toledo, Cheng Fang, Brian Williams. **PARIS: a Polynomial-Time, Risk-Sensitive Scheduling Algorithm for Probabilistic Simple Temporal Networks with Uncertainty**. *Proceedings of the 26th International Conference on Automated Planning and Scheduling (ICAPS-16)*, AAAI Press, London (UK), June 2016. CORE A*.

Xin Huang, Ashkan Jasour, Matthew Deyo, Andreas Hoffmann, Brian Williams. **Hybrid Risk-Aware Conditional Planning with Applications in Autonomous Vehicles**. *Proceedings of the 57th IEEE Conference on Decision and Control (CDC-18)*, Miami (FL), USA, December 2018. CORE A.

Appendix F

Selection of Other Papers

Papers:

Felipe Trevizan, Florent Teichteil-Königsbuch, and Sylvie Thiébaux. **Efficient Solutions for Stochastic Shortest Path Problems with Dead Ends**. *Proceedings of the 33rd International Conference on Uncertainty in Artificial Intelligence (UAI-17)*, Sydney (Australia), August 2017. CORE A*.

Peter Baumgartner, Sylvie Thiébaux, Felipe Trevizan. **Tableaux for Policy Synthesis for MDPs with PCTL* Constraints**. *Proceedings of the 26th International Conference on Automated Reasoning with Analytic Tableaux and Related Methods (Tableaux-17)*, Brasilia (Brazil), September 2017. CORE A.