Technical Report

O-Plan Evaluation Methodology and Experiments

Brian Drabble & Austin Tate

Artificial Intelligence Applications Institute
University of Edinburgh
80 South Bridge
Edinburgh EH1 1HN
United Kingdom

January 6, 1997

ARPA-RL/O-Plan/TR/27 Version 1

Contents

1 Introduction

The aim of this document is to describe the experiments which will be conducted to evaluate features of the O-Plan system. Each experiment aims to show how the planner technologies within the O-Plan system support one or more domain requirements. The generic domain features will be derived from the target domains of the O-Plan system, i.e. the Pacifica NEO, the Atlantica ACP and related materials from the U.S. military, e.g. the JOPES domain evaluation criteria. This information is being used to create an evaluation matrix where the generic domain features form the columns and the generic planning technology features form the rows. For example, one of the domain features from the JOPES evaluation criteria is "plan/concept robustness" which could be required in a number of different domains. This feature can be partially handled in O-Plan by ensuring that the Question Answering (QA) mechanism satisfies world state conditions with more than one contributor. An experiment could be conducted to ensure that the planning technology does indeed provided a way of handling the domain feature. In some cases experiments show how well the requirement is satisfied or show the efficiency of the technology feature in performing the required functionality. In a small number of cases the domain feature will be satisfied by a single technology but in the general case several technologies will need to be combined to satisfy the domain requirement.

This document contains the following information:

- Definition of the generic terminology for domain features
- Definition of the generic planning technology features
- Description (and in later versions of the document the results) of cell experiments which show that a given generic domain requirements can be satisfied by certain technology features of the O-Plan system.

The document is intended to evolve as new domain features and planner technology features are identified and incorporated into the O-Plan system and experience and results are gained from different cell experiments.

The structure of the reminder of the document is as follows: Section 2 describes the project evaluation methodology and background research which has been conducted to date. Section 3 describes the abstract hierarchy of generic domain features. Section 4 describes the abstract hierarchy of generic planning features and Section 5 describes the experiments to be conducted and their results.

A series of appendices are given which provide additional information concerning the generic domain features and planning technologies which have been identified to date. Appendix A describes examples of domain features from the JOPES domain evaluation criteria and categorises them according to the abstraction hierarchy defined in Section 3. Appendix B1 describes the plan representation related domain features of Section 4 categorised according to the <I-N-OVA> constraint model of plans which is being used within the O-Plan project to model plans and planning domains. The aim of the categorisation is to show the domain features identified

can be categorised in the <I-N-OVA> model and to identify possible gaps in the <I-N-OVA> model. Appendix B2 describes the generic planning reasoning features identified in a number of planning systems. Detail provided in this Appendix deepens those found in Section 4 and identifies the systems and technologies which gave rise to the feature. Appendix C relates the domain features described in Appendix A to the technology features described in Appendices B1 and B2. This categorisation has been used to create the initial evaluation matrix and will now form a basis for the design of an initial set of evaluation experiments. Appendix D describes the domain experiments conducted in more detail than that found in Section 5.

2 Project Evaluation Methodology

The O-Plan project includes an ARPI programme-related work package to develop a graded series of demonstrations to show the value of the emerging concepts and technology in a realistic domain setting. This will act as a focus for the work. However, a more technical evaluation of the research is also planned to compliment these demonstrations. This will focus around a selective set of scaled experiments to partially populate an evaluation matrix as follows:

- the columns will be labelled using, as a base, the recent work between AIAI, ISI and ISX to understand the main elements of evaluation for plans in the NEO domain (with later extension to the Air Campaign Planning domain). Every attempt will be made to make this characterisation in generic terms. An example of a column would be "shared use of actions to satisfy (all or some) requirements" which would relate to the NEO "economy of force" element of evaluation.
- the rows will be labelled with technology features of O-Plan related to the emerging terminology from several areas:
  1. Kambhampati's Refine-Plan framework (termination criteria, PMO selection method, bookkeeping method and tractability refinement methods) [8].
  2. Extensions to this framework for Hierarchical Task Network (HTN) planning [1],[9], [10],[19].
  3. Extensions to this framework to allow an explicit agenda of outstanding issues in the plan
     [11], [15], [16].
  4. The emerging KADS model of planning as a generic task [18] using Plan and World Description features, Function Structure, Control Structure, etc..
  5. Gil and Linster's planning application analysis framework [4].
  6. Tate's "standard" components and Plan Modification Operator (PMO) cyclic processing abstraction [14] and the terminology of the <I-N-OVA> constraint model of plans [15].
- each cell then becomes a candidate for comparative evaluation of any system containing the technology feature, or to study the value of any system purporting to address the domain-related feature.

Methods by which experiments will be proposed to populate each cell include:

- use of Kambhampati's "design tradeoffs" [8] approach to characterise the domain assumptions under which the technology features are and are not valuable in domain terms.
- empirical study of the performance of O-Plan using these insights against a realistic evaluation suite for some chosen part of the overall matrix to see whether the predicted influences do occur in the domains of interest.

- use of plan evaluation criteria relevant to the application domain [3],[6].
- insights gained from work with the University of Massachusetts on TIE #5.

Examples of well defined evaluations within specific cells are the Ph.D. projects and ARPI work of Pollack and her colleagues at Pittsburgh. Joslin and Pollack [7] empirically studied the merits of using an opportunistic estimator to prioritise the order in which issues (flaws) are processed in a planner. This work was based on isolating the benefits gained by using the Branch-1 estimator heuristic used in O-Plan [2]. Young and Pollack [20] studied the value of delaying certain types of condition satisfaction during planning. This work relates to the value of the `unsupervised` delayed evaluation condition type in Nonlin [13] and O-Plan.

We will work with others within and beyond the ARPI to provide a framework for evaluation of the benefits of the variety of innovative planning technologies within O-Plan and to encourage population of the matrix through student projects, Ph.D. work, and other project-related work.

Testing, Evaluation Experimentation and Instrumentation

The O-Plan prototype accommodates a number of facilities to allow for development testing, evaluation experimentation, instrumentation, etc. An auto-testing capability allows a package of domain descriptions and tasks to be provided to the planner, one or more plans to be generated for each task in the selected domains and the results to be compared to previously generated plans. A "sanity checker" for plans automates much of the inspection that would otherwise be needed to check that the plans generated are valid against a set of criteria. These facilities allow for repeatable testing of "hooks" for extending the types of automatic testing that can be done on the plans generated by O-Plan.

The O-Plan Task Formalism (TF) domain description language compiler already provides good levels of diagnostic support to domain writers and will be extended during the proposed project to provide feedback to task assigners. The Task Formalism Compiler can be run separately to the planner to provide such information.

3 Generic Domain Features

The aim of this section is to define the generic domain features used to populate the columns of the evaluation matrix. The initial set of features was provided from the JOPES evaluation criteria used for the PRECIS domain [12]. The planning staff estimation process uses 23 of the 39 JOPES identified elements of evaluation for Non-combatant Evacuation Operations (NEOs). Additional generic domain features have been added from the Air Campaign Planning/Atlantica domain and further domain features will be added as materials from this domain are made available.

Each of the domain features from the different domains has been abstracted to identify the class of information or knowledge it represents. For example, "Are the airfields close to the evacuation areas" can be described as a generic spatial constraint between two specified and known locations. A few examples of the generic feature are shown with the bullet items. The features list will be modified as new domain requirements are identified and other classified as being further instances of an already identified class.

The categories identified are described in a separate section.

3.1 Legislation, Agreements and Delegations

1. Are there formal or informal agreements in place on or between the parties taking part in the plan.
   - Overfly rights – neutral cooperation
   - Non-aggression Treaties – customer and supplier
2. Does the plan require cooperation with a body not under direct control of the planners.
   - local authorities – police and civil forces
3. Have all parts of the planning process been delegated to the appropriate bodies
   - Joint Operations – allocation of tasks and missions

3.2 Spatial

1. Do a specified list of locations meet a specified spatial constraint between them.
   - Numerical measures – minimum distance, maximum elevation, minimum area covered, etc.
   - Qualitative measures – "Are the airfields close to the evacuation areas?".
2. Does a specified list of locations have a specific set of attributes.
   - Resources – "Are there people in Abyss to be evacuated"
   - Spatial – "Is there 25,000 square feet of parking at Delta"

3.3 Resources

1. Does a specified list of resources or resource classes have a specific set of attributes.
   - Numerical measures – size, capacity, length, colour, etc.
   - Qualitative measures – "Are the airfields capable of supporting the proposed aircraft types?"
2. Are the attributes of a resource or class of resources fixed or can they be augmented by planning activities.
   - Fixed – length of runways
   - Variable – "Do the airfields have refueling (only if necessary) or do we need to bring it in?"
3. Are the resources and their attributes time dependent and can they be predicted.
   - Unpredictable – "Are we prepared for contingencies with respect to the needed ammunition?".
   - Predictable – Sanctions and restrictions as the result of the military response.

3.4 Temporal Constraints

1. Are there parts, activities or phases of the plan which are time dependent.
   - Point constraint – D-Day, 6th June 1944, etc.
   - Range constraint – Activity 3.1.1. must be accomplished at night, sometime within the next 18 hours.

3.5 External Factors

1. Are there external environmental factors which will influence (positively or negatively) the plan.
   - Predictable: Tide times, phases of the moon, etc
   - Unpredictable: Weather, etc
2. Are there third parties who activities may influence the plan
   - Postive: Suppliers re-equipping their plant
   - Negative: Activities undertake by an enemy to undermine our operations, business competitors launching a new product.

3.6 Solution Quality

1. Does the plan have a specific set of domain attributes:
   - Numerical: Resource Utilisation less than 80% of maximum, contains less that 100 activities.
   - Qualitative: "Is the concept/plan robust? (no/minimal single point failures)", "Is the concept flexible (is this option able to adapt to worsening/improving conditions)? "
2. Is the plan clear in its proposed solution.
   - Phases – "Commence the air operations only"
   - Planning Levels – "Develop the COA to the CONPLAN level"
   - Termination Criteria – "Cease offensive operations when this level of target is reached"
3. Does the plan have a specific set of solution attributes:
   - Numerical: the plan was generated in less than 10 seconds
   - Qualitative: the plan was generated in the minimum amount of time.

3.7 Knowledge Availability

1. Can we acquire the necessary information about a plan attribute at a specified time.
   - Location: "What are the enemies' current intentions"
   - Resource: "What resources are available on D-Day+3 at location 321"
2. Can we acquire the necessary information about a plan attribute at specific intervals.
   - Location: "Provide enemy intentions every 30 minutes"
   - Resource: "Provide aircraft availability lists every 12 hours"

4 Generic Planning Technology Features

The aim of this section is to define the planning technology features used to populate the rows of evaluation matrix. The initial set was defined by examining the features found in planning systems such as O-Plan, SIPE, DITOPS, etc and through studying the current state of the art in planning system technologies [8], [5]. Examples if the features are given in the bullet items. The features can be divided into two main categories: plan representations and planning reasoning.

Plan Representation Features

4.1 Plan State

1. The basic mechanism used to represent the activities in the plan.
   - world state: each state represents a model of the world
   - plan state: each state represents a partial plan
2. The plan state has the ability to handle alternative plan states.

4.2 Abstraction

1. The plan state has the capability to represent abstraction
   - Operator abstraction: decomposition of operators
   - Goal abstraction: goal ordering

4.3 Hierarchies

1. The plan state has the capability to represent explicit hierarchies of activities
   - Action decomposition: activity decomposition

4.4 Activity Ordering Mechanism

1. The mechanism used to order the activities of the plan.
   - totally: the activities of the plan are totally ordered.
   - partially: the activities of the plan are partially ordered by a simple `before` relationship.
   - flexibly: the activities of the plan are ordered by a richer temporal representation.

4.5 Constraint Handling Mechanisms

1. The plan state has the capability to represent constraints on plan entities.
    - Separate constraint managers: A separate constraint manager is used for each major class of constraint represented in the plan state:
        - causality ordering
        - resources
        - temporal ordering
        - spatial
        - domain objects
    - Single constraint set: all constraints of the plan state are held a single set and solved as a single problem.

Planning Reasoning Features

4.6 Planning Status

1. The planning system maintains an explicit list of the problem outstanding in a plan.
    - Agenda/Flaw information: a particular node needs to be expanded
    - Goal information: a given goal needs to be dealt with in the plan.

4.7 Plan Modification Method

1. The mechanism used to modify plan states.
    - re-write rules: theorem prover.
    - procedural code: implicit models of plan modifications.
    - knowledge sources: explicit models of plan modifications.

4.8 Search Mechanism

1. The mechanism used to search for a plan.
    - Breadth first: searching level by level
    - Depth first: searching to a depth cut off
    - Opportunistic: taking the most constrained choice next
2. Attributes of the search mechanism
    - Time: Can a solution be found in a specified time.
    - Quality: Can a solution be found with a specified "quality".

- Probability: Can a solution be found with a specified probability.

3. Book keeping of protected ranges and plan rationale
4. Goal Ordering.
   - linear: The solution assumes multiple goals are independent and can be solved in sequence.
   - non-linear: The solution does not assume multiple goals can be solved in sequence and attempts to interleave the solution of multiple goals.

5 Evaluation Matrix

The aim of this section is describe the evaluation matrix which is used to relate the domain features to the planning technologies identified in Section 3 and Section 4 respectively. The generic domain features will be used to create the columns of the matrix and the planning technologies the rows. The aim of the project is not to conduct every possible experiment but to be selective in its choice experiments. Experiments will be chosen which relate directly to the needs of our TIE partnerships and to the input the project is providing to the series of IFD's (e.g IFD-5 and IFD-6).

Each cell of the matrix will be used to record the outcome of an experiment. An experiment will be in one of three kinds:

- **Validation**:
  A validation experiment will determine whether a specific technique (or series of techniques) can be used to satisfy a domain requirement. In the simplest experiments the result with be either `yes` or `no`. For example, can a spatial reasoning system handle metric distance constraints between two specific points on a map? However, some experiments may provide a qualitative measure of how well the technique handled the domain requirement.

- **Comparison**:
  A comparison experiment will show how one technique performs in relationship than another in handling a specific domain requirement. For example, fixed non-sharable resources can be handled using typed world state conditions or by specific reasoning about the allocation and deallocation of resources. An experiment could seek to determine which technique is better and where appropriate describe the quantitative or qualitative advantage it has.

- **Scalar**:
  A scalar experiment will determine the size and complexity of a problem the technique is capable of handling. In the simplest experiments the result with be a numerical value above which the techniques fails to cope, e.g. when the number of fixed unit resources exceeds 200 typed world state conditions may fail to find a solution. However, some experiments may provide more detailed measures of the complexity of the problem e.g. number of constraints involved before the technique becomes unusable. For example, the technique may fail with far fewer fixed unit resources if there are greater than 20% involved in more complex constraints.

A number of simple experiments have been conducted to validate the approach and the methodology we are using. The list of experiments described in this section will increase as more experiments are conducted. Here the list should be considered as an illustration of the proposed cell experiment design method.

5.1 Experiment 1: Use of Branch 1/Branch N estimators to guide issue selection

Type: Validation

Domain Columns Addressed: 3.6.3
The need for the planner to create solutions in the minimum amount of time.

Technology Rows Addressed: 4.8.1. 4.8.2
The aim of the experiment was to validate that Branch1/Branch N estimators provided support to identify the most constrained issues in the plan state.

Assumptions and Background:
Experimental evidence suggests that least cost flaw repair (LCFR) techniques improve on other issue selection techniques in enabling larger and more complex problems to be solved [7].

Description of Experiment:
The experiment was to introduce branch1/branchN estimator field into each agenda entry. O-Plan's controller used this estimator first to decide which agenda entries were most constrained. Further domain information e.g. level information was used to resolve tie breaks between agenda entries with the same branch1/branchN estimators. The suite of test problems provided with the O-Plan release system was used for this experiment.

Summary of Result:
The outcome of the experiment was that LCFS techniques do provide a way of minimising solution time under certain circumstances. There is evidence to suggest that certain domains in which there are only ever single choices may cause the planner to deep dive in its search for a solution – getting stuck a poor search branches.

Reference to Results:
No further details of the experiment are provided in this version of the document.

5.2 Experiment 2: Use of Level information to guide issue selection

type: Validation

Domain Columns Addressed: 3.6.3
The need for the planner to create solutions in the minimum amount of time.

Technology Rows Addressed: 4.8.1
The aim of the experiment was to validate that taking level information into account when triggering issues on the agenda would avoid the problem of the planner having to back track because it had committed to early to a choice of contributor to satisfy a condition.

Assumptions and Background:
Experimental evidence [17] suggests that delaying certain planning issues until all potential contributors and deletors have been added to the plan enables large problems to be solved

quicker. The hypothesis is that this technique avoids the planner committing early to a potential contributor and then needing to backtrack later when the choice becomes invalidated.

Description of Experiment:
The Task formalism compiler of the O-Plan system was modified to generate level information concerning world state conditions and effect. This was used to validate the domain description to ensure no domain level coding errors were present. O-Plan's controller was modified to hold back agenda entries which required further potential contributors and deletors to be added to the plan state. The suite of test problems provided with the O-Plan release system was used for this experiment.

Summary of Result:
The outcome of the experiment was that level information was useful and allowed a number of planning issues e.g. action expansion, condition satisfaction, etc to be delayed until a their correct time. In a number of cases there seemed to be conflict between the advice from the Branch1/Branch N estimators and that from the level information. This needs to be investigated further.

Reference to Results:
No further details of the experiment are provided in this version of the document.

References

[1] Barrett, A. and Weld, D.S., Task-Decomposition via Plan Parsing, Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94), Seattle, USA, 1994.

[2] Currie, K.W. and Tate, A., O-Plan: the Open Planning Architecture, Artificial Intelligence, 51(1), Autumn 1991, North-Holland.

[3] Gil, Y., Tate, A. and Hoffman, M., Domain-Specific Criteria to Direct and Evaluate Planning Systems, Proceedings of the ARPA/Rome Laboratory Planning Initiative Workshop, (ed. M.Burstein), Morgan-Kaufmann, 1994.

[4] Gil, Y. and Linster, M., On Analyzing Planning Applications, in preparation, 1994.

[5] Hertzberg, J., On Building a Planning Toolbox, Proceedings of the Third European Workshop on Planning (EWSP-95), Assisi, Italy, September, 27-29, 1995 pp3-18. In New Directions in Planning (eds. Gallib, M. and Milani, A.), Frontiers in AI and Applications Series, No 31, IOS Press, Amsterdam, 1995.

[6] Hamazaki, T., High Quality Production Scheduling System, Proceedings of the First Singapore Conference on Intelligent Systems, Singapore, September 1992.

[7] Joslin, D. and Pollack, M.E., Least-Cost Flaw Repair: A Plan Refinement Strategy for Partial-Order Planning, Proceedings of the National Conference on Artificial Intelligence (AAAI-94), Seattle, WA, USA, 1994.

[8] Kambhampati, S., Design Tradeoffs in Partial Order Planning, Proceedings of the Second International Conference on AI Planning Systems (AIPS-94), Chicago, IL., USA, 1994.

[9] Kambhampati, S., Comparing Partial Order Planning and Task Reduction Planning: a Preliminary Report, Proceedings of the Workshop on Comparative Analysis of AI Planning Systems, AAAI-94, Seattle, USA, 1994.

[10] Kambhampati, S. and Hendler, J., A Validation-Structure-Based Theory of Plan Modification and Reuse, Artificial Intelligence, May, 1992.

[11] Penberthy, J. and Weld, D., UCPOP: A Sound, Complete, Partial Order Planner for ADL, Proceedings of the Third International Conference on Knowledge Representation and Reasoning, 1990.

[12] Reece, G., Tate, A., Brown, D., and Hoffman, M. *The PRECiS Environment*, papers of the ARPA/Rome Laboratory Planning Initiative Workshop at the National Conference on Artificial Intelligence (AAAI-93), Washington D.C., USA.

[13] Tate, A., Generating Project Networks, Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-77), Cambridge, Mass., USA, 1977.

[14] Tate, A., The Emergence of "Standard" Planning and Scheduling System Components, in Current Trends in AI Planning (eds. C.Backstrom and E.Sandewall) IOS Press, Sweden, 1993.

[15] Tate, A., Representing Plans as a Set of Constraints - the <I-N-OVA> Model, to appear in the Special Issue on "Evaluation of Plans, Planners, and Planning Agents", ACM SIGART Bulletin, January 1995.

[16] Tate, A., Drabble, B. and Kirby, R., O-Plan2: an Open Architecture for Command, Planning and Control, in Intelligent Scheduling, (eds, M.Zweben and M.S.Fox), Morgan Kaufmann, 1994.

[17] Tate, A., Drabble, B. and Dalton, J., The use of Condition Types to Restrict Search in an AI Planner, Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94) Seattle, USA, August 1994.

[18] Valente, A., Knowledge-Level Analysis of Planning Systems, Proceedings of the Workshop on Comparative Analysis of AI Planning Systems, AAAI-94, Seattle, USA, 1994.

[19] Yang, Q., Formalizing Planning Knowledge for Hierarchical Planning, Computational Intelligence, Vol. 6, No. 1, pp. 12-24, 1990.

[20] Young, R.M., Pollack, M.E. and Moore, J.D., Decomposition and Causality in Partial-Order Planning, Proceedings of the Second International Conference on AI Planning Systems (AIPS-94), Chicago, IL, USA, 1994.

Appendix A: Examples of Domain Features

The aim of this appendix is to describe each of the JOPES domain evaluation criteria in terms of the domain features identified in section 3. In each case the JOPES domain evaluation criteria could be categorised in each of the major catergories defined in section 3.

- **Agreements, Laws, Treaties and Delegations**

  1. A route exists between two specified locations. For example, "Do we have overfly rights and freedom of navigation for all lines of communication?"
  2. Are access rights available to points along a specified route. For example, "Do we have basing rights for all staging bases, intermediate locations and safe havens?"
  3. Do all interested parties and partners agree with the course of action. For example, "Do we have all necessary host nation support at each location? (NB. This can be moral and material)."
  4. Have all parts of the planning process been delegated to the appropriate bodies, For example, "Is this a joint operation? If so have all tasks/missions been allocated?".
  5. Does the proposed plan break any existing Agreements and Treaties. For example, "Is the concept of operations in accordance with all guidance and constraints currently applied?"
  6. Does the proposed plan break any local laws in the area. For example, "Would we be violating any local or international Laws or treaties in conducting these operations?"
  7. Does the plan require cooperation with a body not under direct control of the planners. For example, "Will we be coordinating with local peacekeeping authorities"
  8. Are there policies and guidelines which must be adhered to during specified time periods. For example "Will the operation be able to be conducted within the specified rules of engagement?".

- **Spatial**

  1. Do a specified list of locations meet a specified spatial constraint i.e. minimum distance, maximum elevation, minimum area covered, etc. For example, "Are the airfields close to the evacuation areas?".
  2. Does a specified location provide a large number of resources to an aggressor. For example, "Are the enemy facilities a centre of gravity for their operations? Can they be disabled?".
  3. Does the geography of an area limit the type of plan and the resources which can be used. For example, "Does the terrain/geography inhibit or facilitate the operation?"

- **Resources**

  1. Does a specified resource or class or resources have a specific attribute i.e. size, capacity, length, colour, etc. For example, "Are the airfields capable of supporting the proposed aircraft types?"

2. Are there specific quantities or types of a resource or class of resource available at a specified location. For example, "Are there enough of the right types of staff (refuelers, ATC, maintenance, etc)?"
3. Can the attributes of a resource be augmented by planning activities. For example, "Do the airfields have refueling (only if necessary) or do we need to bring it in?"
4. Are there resources in the area which must be removed or denied access by a third party. For example "Are their American firms which will require staff and essential records/equipment evacuated?"
5. Can a specified type and or quantity of a resource be at a specified location by a specified time. For example, "Can we acquire the ammunition in a timely manner to support operations?"
6. Can we deal with sudden increases in the needed level of a resource. For example "Are we prepared for contingencies with respect to the needed ammunition?".
7. Are there sanctions and restrictions expected to limit the availability of a resource after a certain time point. For example, "Will there be repercussions of the US response e.g. sanctions, boycotts, etc?"
8. Can a specified resource or class of resources be used at a specific location. For example, "Are the friendly forces trained to support operations in this type of area and terrain".

- **Temporal Constraints**

1. Are there parts of the plan which must be achieved by a certain date or point. For example, "Are the necessary logistics operations completed by D-Day".
2. Are there parts of the plan which must be achieved during certain time intervals. For example, "Can critical parts of the mission be done at night?".

- **External Factors**

1. Are there external factors which will influence the plan. For example, "Will weather potentially hamper or delay operations?, Do the tides negatively affect the operation".
2. Can external factors be used to increase the success of the plan. For example, "Can the weather be used to hamper or delay enemy activities or reactions, "
3. Are their known responses which could be used to block the plan. For example, "What activities might the enemy undertake to undermine our operations?".

- **Solution Quality**

1. Is the plan robust with respect to expected changes in the domain and task. For example, "Is the concept robust? (no/minimal single point failures)".
2. Is the plan adaptable to change in the domain or task requirement. For example, "Is the concept flexible (is this option able to adapt to worsening/improving conditions)? "

3. Is the plan clear in its proposed solution, For example, "Are the success, termination and transition criteria well defined'.

- **Knowledge Availability**

1. Can we acquire the necessary information about a location at a specified time. For example, "Do we have sufficient information about the local geography and topology".
2. Can we acquire the necessary information at specified intervals. For example, "Can we get information regarding the agencies, facilities, and resources involved and updates on the status over the course of the operation"

Appendix B1: Examples of Plan Representations – <I-N-OVA> Framework

Each of the generic planning technology features that relate to plan representation described in Section 4 has been classified into a hierarchy defined using the <I-N-OVA> constraint model of plans [15]. Associated with each plan representation feature is the name of a system or piece of research which gave rise to it. This classification hierarchy will be used to create the plan representation related rows of the evaluation matrix.

To validate this proposed classification it has been used to relate the plan representation technology features against the domain features of the JOPES evaluation criteria. Details of the classification can be found in Appendix C.

- **Issues**
  - Planning Status: *agenda* – O-Plan
  - Planning Status: *goal information/stack* – Strips
- **Nodes/Activities**
  - Activity Ordering Mechanism:partially *include nodes* – O-Plan
  - Activity Ordering Mechanism:flexibly *exclude nodes* – Vicar
  - Activity Ordering Mechanism:Flexibly *disjunctive nodes* – Nonlin (Secker), Warplan-C
- **Orderings**
  - Activity Ordering Mechanism:partially *relative*
    - *explicit user directed* – Nonlin, Sipe
    - *implicit through conflict resolution* – NOAH, O-Plan
      - *linear* – UCPOP
      - *partial* – O-Plan, Sipe, Nonlin
  - Activity Ordering Mechanism:flexibly *numerical* – Deviser
- **Variables**
  - Constraint Handling:objects *schema*
    - *type* – O-Plan
    - *codesignation* – Nonlin
    - *non-codesignation* – Nonlin
  - Constraint Handling Mechanism:objects *plan state*
  - Constraint Handling Mechanism:objects *type* – O-Plan, Sipe
  - Constraint Handling:Mechanism:objects *codesignation* – O-Plan
  - Constraint Handling:Mechanism:objects *non-codesignation* – O-Plan

- **Auxiliaries**
  - Constraint Handling Mechanism *resources*
    - *consumable* – O-Plan, Sipe, CAMPS, AMPS
      - *strictly*
      - *producible/by-agent*
      - *producible/outwith-agent*
      - *producible/by and outwith-agent*
    - *reusable* – O-Plan, Sipe
      - *non-sharable*
      - *sharable-independently*
      - *sharable-synchronously*
  - Constraint Handling Mechanism *authority* – O-Plan
  - Constraint Handling Mechanism *spatial*
    - *compute condition* – Nonlin
    - *map* – shaky
  - Constraint Handling Mechanism *causality*
    - *dynamic*
      - *process theory* – Excalibur
      - *simulation* – Zeno
    - *static*
      - *point constraints* – Nonlin
      - *range constraints* – Nonlin
      - *typed preconditions* – Nonlin, O-Plan, Sipe
      - *delayed assignments* – UCPOP (Peot and Smith)
      - *persistent/always facts* – Nonlin, O-Plan
      - *add and delete list* – Strips
  - Constraint Handling Mechanism *other*
    - any others not explicitly represented above – O-Plan
    - informal annotations – O-Plan

Appendix B2: Examples of Planning Reasoning Features

The aim of this appendix is to describe the generic planning reasoning features identified in a number of planning systems. The technologies have been divided into a number of major subgroups based on the type of knowledge the technology is using or manipulating.

- *Control Strategies*
  - *issues* – O-Plan
  - *goal stacks* – Strips, Interplan
  - *fixed loop* – Nonlin
- *Plan Assessment*
  - *alternative plan states*
    - *issues remaining* – O-Plan
    - *"distance" to solution* – O-Plan
  - *current plan state*
    - *branching factors* – O-Plan, Descartes
    - *meta-control* – Molgen, CAMPS
    - *compound agendas* – O-Plan
    - *priorities* – O-Plan
    - *termination conditions* – POCL, SNLP, O-Plan
    - *subgoal interactions* – Interplan
    - *subgoal promotion* – Waldinger's System
    - *action promotion* – WARPLAN
  - *plan state editor*
  - *backtracking*
    - *dependency directed*
    - *temporal*
- *Plan Ordering*
  - *least commitment*
    - *depth first* – Strips
    - *breadth first* – Sipe
    - *hierarchical-strict* – Sipe
    - *hierarchical-non-strict* – O-Plan
    - *looping & iteration* – Sipe, O-Plan
  - *linear*
- *Plan Modification Operators*

- *pieces of code* – Nonlin
- *knowledge sources* – O-Plan, HearsayII
- *re-write rules* – theorem provers
- *critics* – Noah

Appendix C: Domain Features Related to Technology Features

The aim of this appendix is to describe the domain features identified in Section 3 in terms of specific representational features of planning systems. For example, "Are the airfields close to the evacuation areas" can be described as a generic spatial constraint which would be used to rule out plans whose evacuation distance was greater than a prespecified limit. The categories and the generic constraints are detailed in the list below.

Those marked with `OR` indicate a disjunction between different types of constraints, e.g. satisfy the constraint using existing resources e.g. schedule an existing delivery earlier or add new actions to the plan to make the resources available e.g. plan to move an extra Engineers Regiment to the combat theatre. Where a subclass of a constraint class has been identified e.g. a resource of type consumable this is indicated by having the parent and subclass separated by a `/` e.g. Resource/Consumable. Where more than one subclass of a constraint is needed the subclasses are marked with a `+`.

Each of the evaluation criteria has been associated with a descriptor describing the type of information the evaluation criteria is attempting to identify.

- Agreements & treaties
  - **Authority**: Do we have overfly rights and freedom of navigation for all lines of communication?
  - **Authority**: Do we have basing rights for all staging bases, intermediate locations and safe havens?
  - **Resource**: Do we have all necessary host nation support at each location? (NB. This can be moral and material).
  - **Authority**: Would we be violating and treaties with any country involved while conducting the proposed activities?
- Airfields and air facilities
  - **Spatial**: Are the airfields close to the evacuation areas?
  - **Resource**: Are the airfields capable of supporting the proposed aircraft types?
  - **Resource/Consumable**: Are the airfields capable of supporting the proposed aircraft quantities?
  - **Resource/Reusable**: Are there enough of the right types of staff (refuelers, ATC, maintenance, etc)?
  - **Resource/Consumable**: Do the airfields have refueling (only if necessary) or do we need to bring it in?
  - **Resources/Reusable**: Are the airfields capable of providing the equipment necessary to support aircraft operations (radios, radars)?
  - **Resources/Reusable**: Do the airfields have maintenance facilities e.g. hangers, stands, etc if maintenance is going to be carried out there?

- Allied and friendly cooperation
  - **Nodes OR Vars**: Is this a joint operation? If so have all tasks/missions been allocated?
  - **Authority**: Do we have the political backing of our friends and allies for the operation?
- American firms overseas
  - **Resources/Fixed**: Are their American firms which will require staff and essential records/equipment evacuated?
- Ammunition
  - **Resources/Consumable**: Do we have access to sufficient quantities?
  - **Resources/Consumable**: Do we have access to sufficient types?
  - **Resources/Consumable Orderings**: Can we acquire the ammunition in a timely manner to support operations?
  - **Resources/Consumable OR Nodes**: Are we prepared for contingencies with respect to the needed? ammunition?
- Communications
  - **Resources/Reusable**: Will the host nation communications be sufficient (phones)?
  - **Resources OR Nodes**: Do we need secure communications? If so can we provide it?
- Concept of operations
  - **No Constraint Violations**: Is the concept of operations in accordance with all guidance and constraints currently applied?
  - **Issues**: Is the concept robust? (no/minimal single point failures)
  - **Multiple Constributors**: Is the concept flexible (is this option able to adapt to worsening/improving conditions)?
  - **Issues**: Are the success, termination and transition criteria well defined
- Effects of US response
  - **World State**: Will there be repercussions of the US response e.g. sanctions, boycotts, etc?
  - **Authority**: Will the American people support the operation?
- Environment & weather
  - **Temporal**: Can critical parts of the mission be done at night?
  - **World State**: Will weather potentially hamper or delay operations?

  - **World State**: Can the weather be used to hamper or delay enemy activities or reactions.
  - **Temporal** Do the tides negatively affect the operation?

- Facilities (US & Allied)
  - **Resources/Reusable**: Are allied and US facilities sufficient to support operations?
  - **Resources/Consumable**: Intermediate locations: food, water, shelter, safety?
  - **Resources/Reusable+Consumable**: Safe havens: food, water, shelter, hospital, political, onward transportation?

- Facilities (enemy)
  - **Spatial OR Nodes**: Are the enemy facilities a centre of gravity for their operations? Can they be disabled?
  - **Nodes**: Can enemy facilities be captured or utilised for our benefits?

- Forces (US & allied)
  - **Resources/Sharable**: Are the forces trained for this type of operation?
  - **Resources/Sharable**: Are their sufficient forces to offset anticipated and contingency enemy reactions?
  - **Temporal**: Can the forces be in position in the timescale identified?
  - **Resources/Reusable+Consumable**: Do the forces have sufficient equipment?
  - **Resources OR Spatial**: Can we accomplish the mission with a minimum footprint? (minimum troops, casualties, destruction)

- Forces (enemy)
  - **Nodes**: Can the enemy forces be countered during the operation to minimise their impact, e.g. loss of life?

- Geography and terrain
  - **Resources/Reusable**: Are the friedly forces trained to support operations in this type of area and terrain.
  - **Spatial**: Does the terrain/geography inhibit or facilitate the operation?
  - **Spatial**: Are the beaches accessible as transportation alternative?

- Legal authorities
  - **Authority**: Would we be violating any local or international Laws or treaties in conducting these operations?
  - **Authority**: Will we be coordinating with local peacekeeping authorities.

- Maps and chart availability

  - **Resources/Consumable**: Do we have sufficient information about the local geography and topology (NB. Information as a consumable resource)?

- Medical services
  - **Resources/Consumable**: Sufficient (in both quantity and type) medical facilities must be provided both en-route and at each safe haven.
  - **Resources/Reusable**: Medical units must be available at each of the evacuation centres in country.

- Non-combatant personnel
  - **Resources/Reusable**: Accommodations (both transportation, food and lodgings) must be made available for all evacuees including both US and other friendly nationals evacuated by the US.

- Operational comparison (US and adversary)
  - **Nodes**: What activities might the enemy undertake to undermine our operations?
  - **Multiple Constributor**: How susceptible is our plan to enemy activities?

- Reconnaissance reporting
  - **Resources/Consumable**: Can we get assessments of the enemy activities for this operation (NB. Information as a resource?)
  - **Resources/Consumable**: Can we get information regarding the agencies, facilities, and resources involved and updates on the status over the course of the operation.

- Rules of engagement (ROE)
  - **All Constraints**: Will the operation be able to be conducted within the specified rules of engagement?

- Seaports and port facilities
  - **Spatial**: Are the seaports close to the evacuation areas?
  - **Resources/Reusable**: Are the seaports capable of supporting the proposed evacuation ship types?
  - **Resources/Reusable**: Are the seaports capable of supporting the proposed evacuation ship quantities?
  - **Resources/Reusable**: Are there enough of the right type of support personnel available (refuelers, STC, maintenance)?
  - **Resources/Reusable**: Do the docks have facilities for refueling (if necessary) or do we need to bring it along?
  - **Resources/Reusable**: Are the docks capable of supporting the equipment necessary to support ship operations (radios?)

- Transportation (local)

- **Resources/Reusable**: Is sufficient local transportation available for transport to assembly areas?
- **Resources/Consumable**: Can transportation been purchased or rented locally as opposed to being provided by the evacuation forces?
- **Spatial**: Are the routes susceptable to enemy action?
- **Nodes**: Can the local lines of communications be protected during use?

Appendix D: Detailed Cell Experiments

Experiment 1: Use of Branch1/Branch Estimators to Guide Issue Selection

No Further Details

Experiment 2: Use of Level Information in Guiding Issue Selection

No Further Details