Interactive Configuration based on Incremental Constraint Satisfaction

E. Gelle and R. Weigel
Swiss Federal Institute of Technology (EPFL)
IN-Ecublens, CH-1015 Lausanne, Switzerland,

E-mail : { *Gelle* / *Weigel* } *@lia.di.epfl.ch*

Abstract

In this paper, we focus on techniques for incremental constraint-based configuration with discrete and continuous variables. We show how to formalize constraint knowledge using compatibility and activity constraints (Mittal 1990) and how this knowledge is used for reasoning within an intelligent CAD system. Most technical (as opposed to spatial) constraint configuration systems nowadays use algorithms for solving discrete problems (Haselboeck 1994) We claim that configuration is both *discrete and continuous* in nature and that new methods for handling both constraints in a unified way must be integrated in configuration systems. Visualization of the globally consistent configuration problem space allows a systematic and exhaustive exploration of the space in an interactive fashion (Haroud 1995).

Knowledge maintenance in configuration systems must be simplified, because configuration knowledge of todays products evolves over the whole product life cycle. The knowledge representation in deductive rule-based systems as often used in intelligent CAD systems on the other hand will always be context dependent; maintenance problems resulting from this context dependency are often insurmountable. We have identified the context independence of constraint-based knowledge representation as an important feature for facilitating the incremental development and maintenance of large evolving knowledge bases (Weigel 1992).

Keywords

Rule-based vs. constraint-based configuration, incremental constraint satisfaction, interactive configuration and design

1 INTRODUCTION

In recent years, manufacturing trends have changed from pure mass-production to a more customer oriented one-of-a-kind production. The main reason for this change is that today's customers have very specific and individual requirements, which can no longer be satisfied by mass-products. The one-of-a-kind production of many consumer and investment products requires powerful modeling techniques and representation methods combined with features which facilitate maintenance and extendability. We claim that the framework of incremental constraint satisfaction offers these features.

Knowledge formalization: The advantage of using constraints to formalize design knowledge is that relations between design parameters can be stated without explicitly mentioning of the context in which these relations hold. In section 2 we will show how context-independent knowledge representation of constraint systems facilitates knowledge engineering and maintenance of configuration system during the whole life-cycle of the product. In section 3 we present our framework for dynamic constraints over discrete and continuous variables. In the framework of incremental constraint satisfaction of Mittal (1990) one can reason about the introduction and retraction of variables respectively constraints during problem solving. This modeling technique is, for reasons of modularity and efficiency, especially useful when large amount of constraints must be handled.

Interactivity: Often configuration systems work in a batch-like manner which means that the customer requirements must all be known a priori and are then fed into the configurator to generate for example the bill-of-material of the product. The interactivity in our system leads the user from a rough to a more detailed specification. There is no need to specify all "input" parameters at once. Furthermore, since we based the reasoning within the system on global consistency of the constraints, we can guarantee that the user cannot move into regions of the search space without solution. Although global consistency is computationally expensive, it is especially useful in interactive systems when working with continuous constraints where an enumeration of the single feasible solutions is no longer possible. In section 4 we will describe using a small example how global consistency is integrated in framework of incremental constraint satisfaction.

2 MAINTAINING CONFIGURATION KNOWLEDGE

Todays products evolve during their whole life-cycle. This implies that new knowledge must be integrated and old knowledge must be removed constantly from the configuration system. By using a small example we will show that building and maintaining a constraint knowledge base is much easier than building and maintaining a rule-base. Our fictive car company decided to develop a new *funcar* variant of its product line. The effects of adding this new knowledge to a rule-base respectively to a constraint-base will be studied and analysed.

Rules are described in the format "IF variable1 = value THEN variable2 = value" and a simple forward chainer will be used for reasoning. Constraints are represented using tables and the search could be done by a standard backtracking algorithm. Rules and constraints are shown in Figure 1. The marketing department of the company decides to introduce a new funcar type its is the task of the knowledge engineer to enter rules

```
R1 IF Package = Deluxe
    and Frame = convertible
    THEN Engine = A

R2 IF Package = Deluxe
    and Frame = hatchback
    THEN Engine = B

R3 IF Package = Std
    and Frame = convertible
    THEN Engine = A

R4 IF Engine = A
    THEN Transmission = manual

R5 IF Engine = B
    THEN Transmission = automatic

R6 IF Type = Sportscar
    THEN Frame = convertible

R7 IF Type = Familycar
    THEN Frame = sedan

R8 IF Type = Sportscar
    THEN Transmission = manual
```

Package	Frame	Engine
Deluxe	convertible	A
Deluxe	hatchback	B
..	..	

Engine	Transmission
A	manual
B	automatic
A	half-automatic
..	..

Type	Frame
Sportscar	convertible
Familycar	sedan
..	..

Type	Transmission
Sportscar	manual
Familycar	half-automatic
..	..

Figure 1 *Left:* Rules for Car Configuration. *Right:* Constraints for Car Configuration

```
R10 IF Type = Funcar
     THEN Frame = convertible

R11 IF Type = Funcar
     THEN Transmission = half-automatic
```

Figure 2 New rules to be added to the rule base

R11 and R12 shown in Figure 2 into the rule-base. Simply adding these two rules will render the rule base inconsistent. This can be seen when configuring a *funcar deluxe*. The rule sequence *R10, R1, R4* and *R11* leads to the conflict that the transmission should be manual and half-automatic at the same time. Therefore one needs to modify the rule-base as shown in Figure3 left.

Comparison: Constraints and rules must be interpreted differently. Consider for example the allowed tuple (A manual) in the constraint between engine and transmission. The constraint must be interpreted as follows: "engine A is compatible with manual transmission" while the interpretation of rule 4 is "every car with engine A will get a manual transmission". The scope of the constraints is local in the sense that new knowledge about funcars for example can not invalidate the constraint knowledge. The scope of the deductive rule on the other hand is global and new knowledge can invalidate the rule as described above.

In systems built using deductive rules, in particular expert systems, the context-dependence results in severe problems of *maintenance* of knowledge in the face of a dynamic world. Even minute changes of technology or changes in the marketing policy require revision of the entire rule set, which can be very costly. In the rule-based approach, adding a new car-

```
Step 1 removing Rule 4
    IF Engine = A
    THEN
    Transmission = manual

Step 2 adding Rule 4a
    IF Engine = A and
    Type = Funcar
    THEN
    Transmission = half-automatic

Step 3 adding Rule 4b
    IF Engine = A and
    Type = Sportcar or Familycar
    THEN
    Transmission = manual
```

Package	Frame	Engine
Deluxe	convertible	A
Deluxe	hatchback	B
..	..	

Engine	Transmission
A	manual
B	automatic
A	half-automatic
..	..

Type	Frame
Sportscar	convertible
Familycar	sedan
Funcar	convertible
..	..

Type	Transmission
Sportscar	manual
Familycar	half-automatic
Funcar	half-automatic
..	..

Figure 3 *Left:* steps to make the rule base consistent. *Right:* Constraints for the extended Car Configuration

type forced us to create a new relation between engine, type and transmission. Expressing knowledge without looking at the context is one of the major advantages of constraint based reasoning. No relation between engine, type and transmission is necessary.

One must furthermore realize, that assigning a single source (e.g. marketing regulations) to the rules 4a and 4b is not possible. Rules from different sources * are mixed together in new rules only to get a consistent chaining behaviour. These new rules can be considered as "interface rules" between knowledge sources and they are artificial in the sense that only the chaining behaviour is responsible for their existence. Proving the correctness of those rules becomes cumbersome. The problem of systems built using deductive rules is therefore not only the revision of the rules, which can be very costly. The revisions itself will make the maintenance of the rule-base even more difficult, leading to systems which are no longer manageable.

Mechanising the knowledge engineering process is straightforward within the constraint-based framework, because whenever a relation between variables is "established", the knowledge engineer must enter all valid variable-value combinations for that relation. Knowledge engineering in the rule-based framework on the other hand resembles a more hand-crafted approach, since it is possible to delay the engineering of rules until they are needed for a specific configuration. Consider the relation between engine and transmission in the above examples, where the tuple (A half-automatic) in the constraint-based system was "valid" from the very first moment. In the rule-based system on the other hand one could find this piece of knowledge only in "decoded" form within rule 4b, which was added after a contradiction was detected.

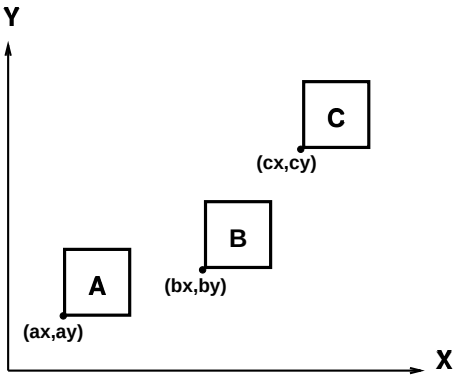
*The knowledge of rules 4a and 4b stems from the marketing **and** the engineering department!

3 INTERACTIVITY AND SOLUTION SPACES

Traditionally, configuration tasks have been reduced to the activity of assembling components of predefined dimensions (Haselboeck 1994) The number of ways such components can be combined is enumerable. The configuration task is thus modeled as a constraint satisfaction problem on finite, discrete variables. Most practical tasks, however, include objects without predefined ranges of dimensions and continuous variables are needed to describe their properties. Consider, for example, the spatial configuration of 3 objects A, B and C. The position of the objects are described by continuous variables.

Example: Configuration with continuous variables Let a_i, b_i, c_i for $i = x, y$ be the x respectively y coordinates of the objects A, B and C. We can formulate a set of constraints involving a_i, b_i, c_i.

C1 $a_i, b_i, c_i \in [0, 10]$ for $i = x, y$
C2 $b_x \leq {a_x}^2 + 2$
C3 $b_x \leq 2c_x$
C4 ${a_x}^3 + 4 \leq c_x b_x$
C5 $b_y \leq a_y$
C6 $a_y + {b_y}^2 \leq c_y$
C7 $c_x \leq c_y$
C8 $2{b_y}^2 + 2c_x \leq 5b_x$



In order to solve these equations, engineers currently apply a particular sequence of calculations, but never consider the entire space of solutions. They will first solve subsets of constraints and then try to combine these partial solutions by picking one feasible point in a subregion corresponding to a subset of constraints and checking it against the resting constraints.

In constraint-based systems, consistency algorithms are used in order to refine the possible solution space for each variable. Search then finds single feasible solutions within the refined space. Depending on the structure of the problem, applying a certain degree of consistency results in a globally consistent solution space. *Global consistency* in a constraint network ensures that for each variable a value can be found so that all the constraints are satisfied. Haroud (1995) has developed an algorithm guaranteeing *global consistency* for continuous constraint satisfaction problems (CCSPs).

Figure 4 visualizes the solution space for the constraints **C1**... **C4** of the spatial configuration example. In this algorithm, cubes approximate the region defined by each constraint in the tree-dimensional space. The algorithm calculates consistent solution spaces by combining these regions. Given a feasible partial solution, an extension to a globally consistent solution is guaranteed. Users can interactively restrain the feasible solution space and focus on regions of interest within. It is now possible for them to explore feasible space for preferable solutions. In Figure 4, the solution has been restrained to $6.5 \leq b_x \leq 7.2$. All the values dependent on b_x (a_x and c_x) are recalculated with respect to the new value of b_x.

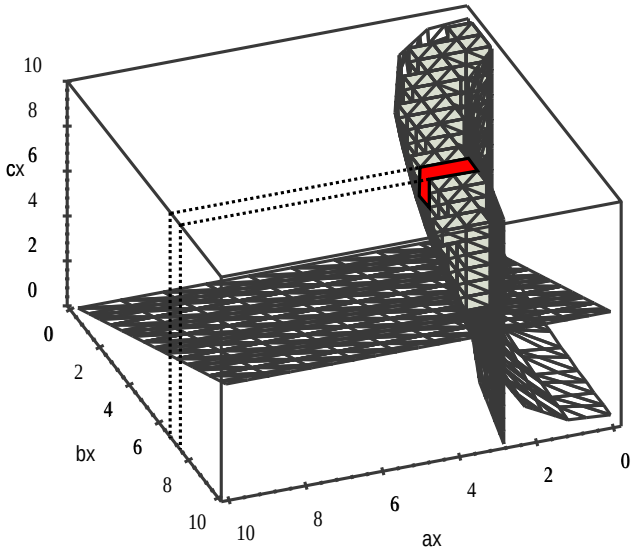


Figure 4 *The solution space corresponding to the constraints* **C1**... **C4**.

4 INCREMENTAL CONSTRAINT SATISFACTION

In design and configuration, the problem space is often huge and interaction between components very complex. To reduce computational complexity, the task is modeled as an incremental constraint satisfaction problem (CSP) defined by a set of variables X, a set of constraints C and a set of initial conditions W. W defines the set of variables that have to be part of every solution. C consists of two types of constraints: *compatibility* and *activity constraints*, noted CC respectively AC. Not all the variables need to be part of a solution: X only defines the space of potentially active variables. It follows, that not all of the constraints will be relevant for one solution. We need to reason on the activity of variables and the constraints depending on them. The introduction of new variables and constraints depends on *activation conditions*. This dependency can be formulated as a so-called activity constraint[†] according to the definitions of Mittal (1990).

AC: $C_i(Y_1) \rightarrow active : Y_2$

This activity constraint activates the subset $Y2$ of variables if the activation condition (or precondition) $C(Y_1)$ is satisfied. The precondition C_i defines a mathematical relation on Y_1. Compatibility constraints define the relations that must hold between active variables:

CC1: $C_i(Y_2)$
CC2: $C_i(Y_3) \rightarrow C_j(Y4)$ $\qquad Y_3 \cap Y_4 = \emptyset$

If all the variables of Y_2 are active in **CC1**, C_i has to be satisfied. The constraint C_i is a mathematical (in)equality on the subset of variables Y_2 or, in the discrete case, a relation between variables where allowed tuples are enumerated. In **CC2**, it depends on the values of Y_3 as well as on the existence of $Y_3 \cap Y_4$ if C_j is relevant or not.

[†] An activity constraint should not be mistaken for a rule, it defines a constraint on the activity of variables in the problem space

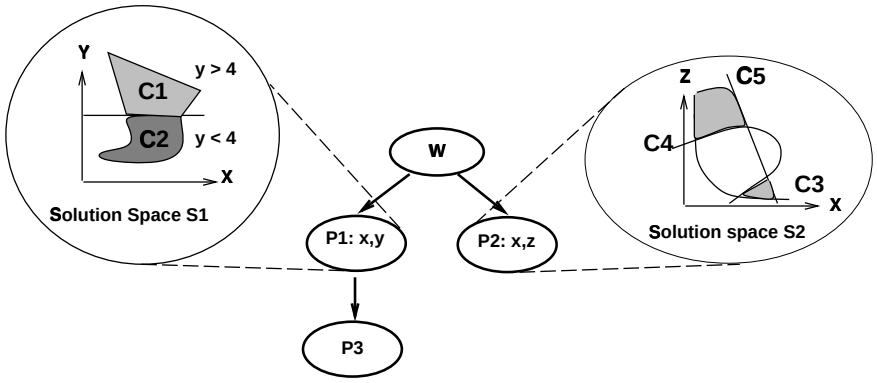


Figure 5 *A sequence of static problems. In problem space P1, the variables x and y are active; in P2, x and z. In the solution space S1, the set of the constraints explicitly depends on the value of y. In S2, the global consistency algorithm finds two separate feasible regions satisfying the same set of constraints with C3:* $y - 1/x \geq 0$, *C4:* $(x-y)^2 + 2x + 2y - 8 \geq 0$, *C5:* $y + 1.3x - 4 \leq 0$.

The goal is to find all the solution spaces S so that for each solution $s \in S$: s satisfies all the constraints defined on a set of active variables in X and no more variables can be activated.

For reasoning on the surface of the objects in the spatial configuration problem, the objects are given additional properties such as shape and dimension. Depending on their shape, the variables width, height or radius are relevant and their surface will be calculated differently. Initially, W is $\{shape_x, surface_x\}$, $x \in \{A, B, C\}$ and the constraint set C is defined by

AC1 $shape_x = rectangle \rightarrow length_x$ AND $width_x$
AC2 $shape_x = circle \rightarrow radius_x$
CC3 $surface_x = radius_x * \pi$
CC4 $surface_x = length_x * width_x$
CC5 $R(shape_A, shape_B) = (rectangle, circle)(circle, rectangle)$

When $shape_A = rectangle$, the variables $surface_A$, $length_A$ and $width_A$ are active and $surface_A$ will be calculated according to **CC4**. $Shape_B$ is restrained to a circle by **CC5** and its surface is calculated according to **CC3**.

The process of finding solution spaces involves an *activate-propagate cycle*: From the given set of active variables, all activity constraints are checked in order to activate new variables. This step defines the new *problem space*, i.e. the space of currently active variables. In the propagate step, the compatibility constraints defined on active variables are checked for global consistency. Feasible partial *solution spaces*, i.e. regions in N-space defining value bounds for the variables, are found. At each cycle, the values of currently active variables are either refined or an inconsistency is detected. This solution space is discarded and the algorithm either backtracks to another solution space or to the next problem space not yet treated. It halts when no new problem spaces can be created, i.e. all the problem spaces have been searched and no new variables can be activated. The final solution spaces are those in the leaves of the problem space tree.

An incremental CSP can then be viewed as a sequence of static problems (Figure 5): $P_0, \ldots P_n$ with $P_0 =< X_0, C_0, D >$ and $P_i =< X_i, C_i, D >$ where $C_i = C_{i-1} \pm \{C_j\}$

with $\{C_j\} \subset C$, X is the set of variables and D the variables' domains. As can be seen in figure 5, activity constraints may split up one problem space into several each containing different sets of active variables ($P1$ and $P2$ in Figure 5). Constraints may split a solution space further by creating separate regions of consistent values. In S1, the relevant constraints explicitly depend on values of y. In S2, the intersection of $C3$, $C4$ and $C5$ create two separate feasible regions.

4.1 Example in Bridge Configuration

We would like to show on a small example of bridge configuration how components of the configuration product can be added incrementally. Adding components leads to new design parameters and values that activate new constraints.

The aim in bridge configuration is to find bridge designs that satisfy design specifications as well as building codes and other requirements as described in Haroud and Boulanger (1995). Given the section of the valley in which the bridge has to be built, a set of initial conditions W and a set of constraints C, we want to find all solution spaces. In the following example, the designer already decided on a beam bridge type. The initial conditions W are thus:

Parameter	**Definition**	**Domain**	**Value**
L	length of the valley	real number	200
bridge type	type of the bridge	{beam,cable-stayed,frame,arch}	beam

and the constraint set C is defined by:

AC1 $bridge\ type = beam \rightarrow \left\{ \begin{array}{ll} nb\ of\ spans: & [1,20] \\ span: & \forall_{i=1}^{nb\ of\ spans}\{span_i : [20, L]\} \\ maximal\ span: & [20, L] \end{array} \right\}$

AC2 $maximal\ span > 80 \rightarrow beam\ type : \{variable\ height, constant\ height\}$

CC1 $beam\ type = variable\ height$

CC2 $nb\ of\ spans \geq 4 \rightarrow \forall_{i=1}^{nb\ of\ spans}\{span_i = \frac{L}{nb\ of\ spans}\}$

CC3 $nb\ of\ spans < 4 \rightarrow \left\{ \begin{array}{l} span_1, span_{nb\ of\ spans} < \frac{L}{nb\ of\ spans} - 10 \pm 10\% \\ \forall_{i=2}^{nb\ of\ spans-1}\{span_i > \frac{L}{nb\ of\ spans} + 20 \pm 10\%\} \end{array} \right\}$

CC4 $\sum_{i=1}^{nb\ of\ spans} span_i = L$

CC5 $maximal\ span = max_{i=1...nb\ of\ spans} span_i$

The activity constraints in this example show how new components are added (**AC2**) and how the structure of the artifact is built (**AC1**). We simplified the structure by representing beams and piers by the unique notion of spans[‡].

Starting with $bridge\ type = beam$ and $L = 200$ the **AC1** is activated and the variables $nb\ of\ spans = [1, 10]$, $span = [20, L]$, $maximal\ span = [20, L]$ are added. The constraints **CC2**, **CC3**, **CC4** and **CC5** are propagated. They split the solution space into two spaces S1 and S2: S1 with $nb\ of\ spans < 4$ and S2 with $nb\ of\ spans \geq 4$. In S1, the constraints **CC3**, **CC4** and **CC5** are considered. In S2, **CC2**, **CC4** and **CC5** are propagated. In the second cycle, a new problem space P2 is created by adding the variable *beam type*

[‡]A span is the distance between two adjacent piers.

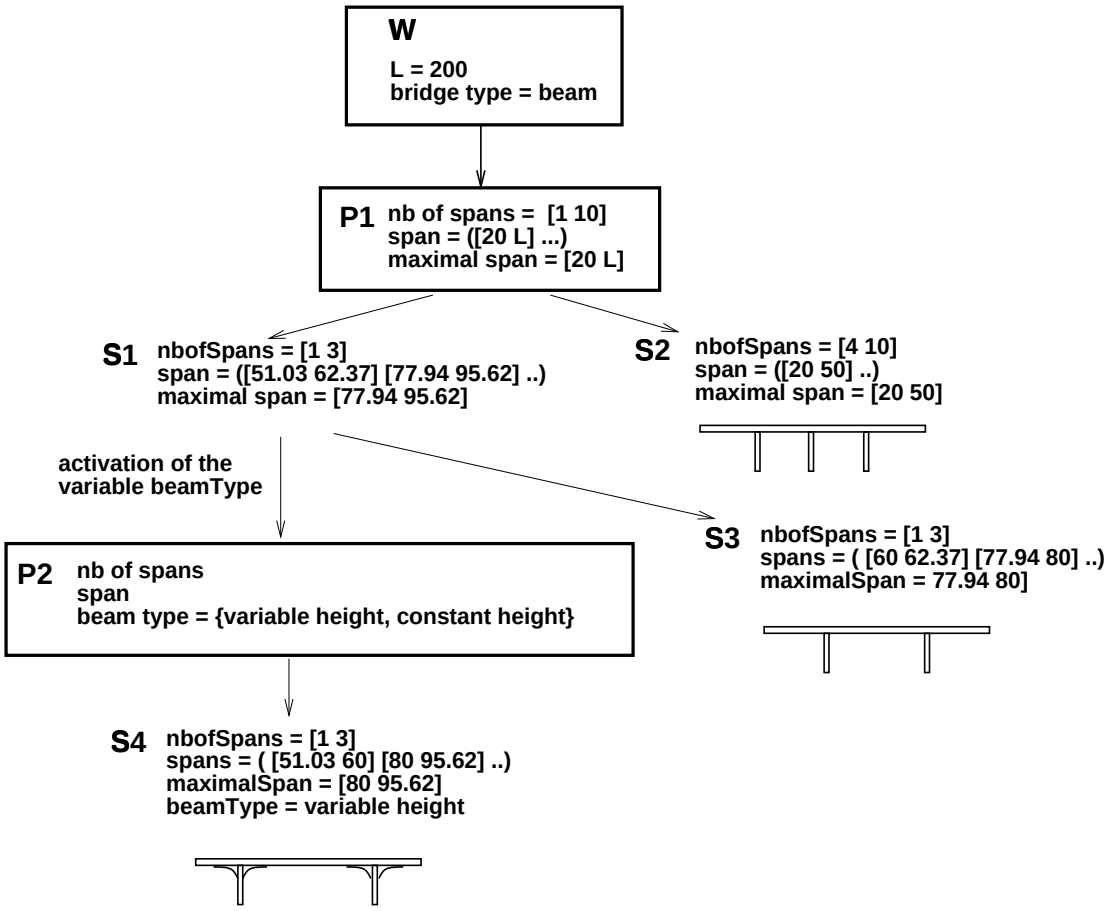


Figure 6 *A preliminary bridge design with two problem spaces and different solution spaces.*

according to the activity constraint **AC2**. Within P2, the constraint **CC1** is propagated adding a new component to the bridge.

This is an example of how configuration can be formalised as an incremental process of adding new components and checking relevant constraints. Inconsistencies and splits in the solution space are detected during constraint propagation (solution spaces in P1). After each constraint propagation the user has the possibility of interactively restricting values. In S1, for example, the user could set *nb of spans* to 3.

Our implementation is based on a forward chaining rule engine for activating constraints, a justification-based truth maintenance system (JTMS) and currently a low-level constraint satisfaction algorithm for checking consistency. During constraint propagation, new feasible regions inferred are justified by a JTMS-label linking design variables and constraints. Each constraint has a JTMS-label as well. Reasoning on the relevancy of a constraint can so be made explicit. After each cycle, the partial result is visualized in ICAD, an intelligent CAD system. It provides the user-interface with graphical representation and a product model of the bridge. New components are incrementally added in the ICAD structure corresponding to our algorithm.

5 CONCLUSION

We have shown how configuration problems can be modeled in the framework of incremental constraint satisfaction problems with discrete and continuous variables. This is an extension of the purely discrete framework of Mittal(1990) and allows us to attack a broad range of configuration problems, which couldn't be solved with discrete variables only. Furthermore, we can guarantee maintenance and extendability of the system within this framework. We have based reasoning on algorithms of global consistency in the constraint network. Although this is computationally expensive, this gives the users the possibility to concentrate on the "solution spaces" of the problems P_i after each propagation step. They are able to explore different partial solutions instead of searching in regions where no solutions can be found. The integration of algorithms for incremental constraint satisfaction into an intelligent CAD system, like ICAD, has shown to be very promising since the user can interact with the system by working with the graphical representations of the configuration objects instead of manipulating alpha-numeric symbols.

ACKNOWLEDGMENTS

Funding for this research was provided by the Swiss National Science Foundation. The authors would like to thank Boi Faltings, Ian Smith, Djamila Haroud, LIA, EPFL, and Sylvie Boulanger, ICOM, EPFL, for useful discussions.

REFERENCES

Mittal, S. and Falkenhainer, B. (1990) Dynamic Constraint Satisfaction Problems. *Proc. of AAAI-90*, 25–32

Faltings, B. and Weigel, R. (1994) Constraint-based knowledge representation for configuration systems. *Technical Report TR-94/95, LIA-DI, EPFL, Lausanne, CH*

Haroud, D. (1995) Global Consistency for Continuous Constraints. *Phd thesis, LIA-DI, EPFL, Lausanne, CH*

Haroud, D., Boulanger, S., Gelle, E. and Smith, I. (1995) Strategies for Conflict Management in Preliminary Engineering Design. *AIEDAM Special Issue on Conflict Management in Design*, 313–323

Stumptner, M., Haselböck, A. and Friedrich, G. (1994) COCOS, A Tool for Constraint-Based, Dynamic Configuration. *Proc. of the 10th IEEE Conference on AI Applications*, 373–380

Mackworth, A.K. and Freuder, E.C. (1985) The Complexity of some Polynomial Network-Consistency Algorithms for Constraint-Satisfaction Problems. *Artificial Intelligence*, **25**, 65-73

Mohr, R. and Henderson, T (1986) Arc and Path Consistency Revisited. *Artificial Intelligence*, **28**, 225–233