

# CONSTRAINT SATISFACTION

Constraint Satisfaction 2: SAT

*PROP<sub>4</sub>*, Let  $T$  be a constant, the last *PROP* predicate is named ***PROP<sub>i</sub>***, and is defined to be the first of the three predicates ***PROP<sub>41</sub>***, ***PROP<sub>31</sub>*** and ***PROP<sub>0</sub>*** (in this order) which has at least  $T$  variables in its denotational semantics.

Each *PROP* predicate results in a DPL procedure, ***PROP<sub>0</sub>***, ***PROP<sub>0</sub>***, ***PROP<sub>3141</sub>*** and ***PROP<sub>i</sub>***, respectively giving *Sata*, *SatO*, *Sat<sub>3141</sub>* and *Satz*, and ***PROP<sub>i</sub>***, giving *Sat<sub>ij</sub>*. These programs are different only in *PROP* predicate, except *SatQ* which need not count the occurrences of variables.

We run the 12 programs (compiled using gcc with optimization) on a PC with a 133 Mhz Pentium CPU under Linux operating system on a very large sample of random 3-SAT problems generated by using the method of Mitchell et al. [Mitchell et al., 1992]. Given a set  $V$  of  $n$  Boolean variables  $\{x_1, x_2, \dots, x_n\}$ , we randomly generate  $m$  clauses of length 3. Each clause is produced by randomly choosing 3 variables from  $V$  and negating each with probability 0.5. Empirically, when the ratio  $m/n$  is near 4.25 for a 3-SAT formula  $F$ ,  $F$  is unsatisfiable with a probability 0.5 and is the most difficult to solve. We vary  $n$  from 140 variables to 340 variables incrementing by 20, for each  $n$  the ratio clauses-to-variables ( $m/n$ ) is set to 4.0, 4.1, 4.2, 4.25, 4.3, 4.4, 4.5. At each ratio and by each program, if  $n < 280$  then 1000 problems are solved, if  $280 \leq n \leq 300$  then 500 problems are solved, if  $n = 320$  then 300 problems are solved, and if  $n = 340$  then 100 problems are solved. A problem is solved successively by all the 12 DPL procedures before another to ensure the same environment to all programs. Due to the lack of space, we only present the experimental results for the ratio  $m/n = 4.25$  in Figures 4, 5, and 6, where the DPL procedures corresponding to the curves are listed in the same order from top to bottom. The experimental results on the other ratios give exactly the same conclusions.

### 3 A Pure UP Heuristic Versus a Pure Mom's Heuristics: *SatO* vs *Sata*

*SatO* systematically examines all the variables by unit propagation at all nodes, using a pure UP heuristic, while *Sata* does not examine any variable so and employs a pure Mom's heuristic. One might believe that *SatO* would be simply too slow, but it is not the case. *SatO* is much faster than *Sata*. In fact from Figures 4 and 5, all DPL procedures using a UP heuristic in our experimentation are substantially better than *Sata* in terms of search tree size and real run time.

Note that Mom's heuristic used in *Sata* is similar to the so-called two-sided Jeroslow-Wang rule [Hooker and Vinay, 1995], with the only difference that a clause of length  $i$  is counted as 5 clauses of length  $i+1$  instead of 2. Our experiments suggest that 5 is better than 2. 5 is

also similar to the exponential factors in C-SAT [Dubois et al., 1993] where 5.71 ternary clauses are counted as 1 binary clause.

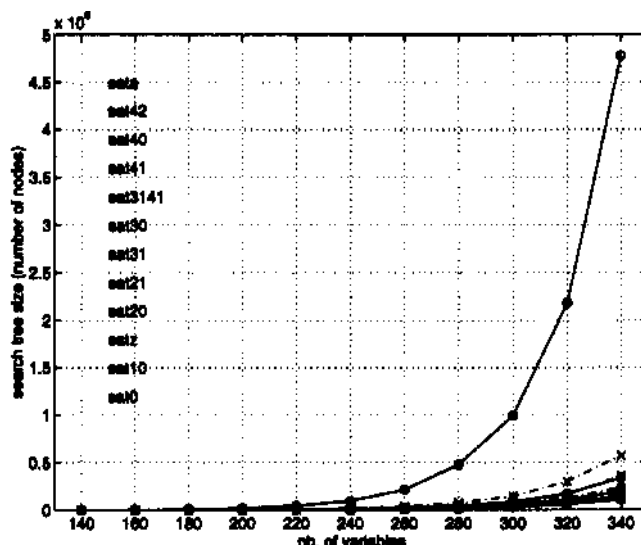


Figure 4: Mean search tree size of each program as a function of  $n$  for hard random 3-SAT problems at the ratio  $m/n = 4.25$

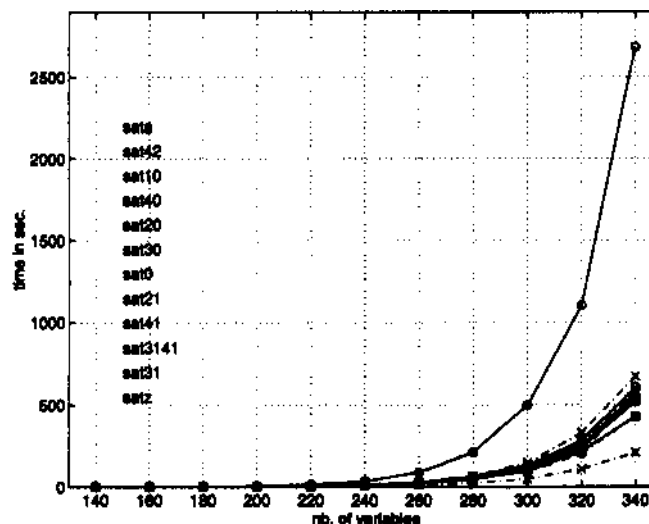


Figure 5: Mean run time of each program as a function of  $n$  for hard random 3-SAT problems at the ratio  $m/n = 4.25$

*SatO* actually is slower than five other programs based on balanced restrictions of variables to be examined by unit propagation, but not substantially so (except *Satz*). The surprisingly good performance of *SatO* confirms the power of UP heuristics for selecting the next branching variable and suggests that its effect for detecting failed literals is only secondary.

### 4 Restricted UP Heuristics

Figure 6 illustrates the number of variables examined by different restricted UP heuristics at a node.

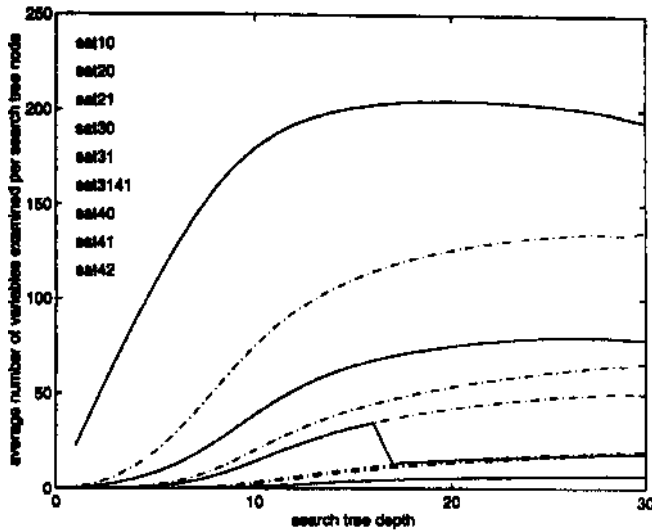


Figure 6: Average number of variables examined at a search tree node in a given depth when solving hard random 3-SAT problems of 300 variables and 1275 clauses (500 problems are solved) for 9 programs

#### 4.1 Restriction by total number of binary occurrences of a variable

Four programs *Sat*<sub>10</sub>, *Sat*<sub>20</sub>, *Sat*<sub>30</sub> and *Sat*<sub>40</sub> realize this type of restrictions. While a classical Mom's heuristic selects the next branching variable having maximum binary occurrences, the restricted UP heuristics examine a set of variables having more binary occurrences than others, including the variable having maximum binary occurrences. From Figure 4, it is clear that the more variables are examined, the smaller the search tree size is.

#### 4.2 Balanced restriction by total number of binary occurrences of a variable

Four programs *Sat*<sub>21</sub>, *Sat*<sub>31</sub>, *Sat*<sub>41</sub> and *Sat*<sub>42</sub> realize this type of restrictions. The *PROP* predicates require that a variable occurs both positively and negatively in binary clauses to balance the search tree. We compare the duet *Sat*<sub>*i*0</sub> and *Sat*<sub>*i*1</sub> (*i*=2, 3, 4) and observe that *Sat*<sub>*i*1</sub> examines strictly fewer variables than *Sat*<sub>*i*0</sub> and is faster than it in spite of a slightly larger search tree. In particular, *Sat*<sub>41</sub> and *Sat*<sub>40</sub> examine almost the same number of variables (see Figure 6), but the balanced restriction gives a faster DPL procedure.

We pay special attention to *PROP*<sub>31</sub> and *PROP*<sub>41</sub> since they seem to be the best balanced restrictions.

#### 4.3 Dynamic restriction as a function of search tree depth

*Sat*<sub>3141</sub> realizes this restriction. A general observation when solving 3-SAT problems using a DPL procedure is that there are more and more binary clauses when descending from the search tree root and the denota-

tional semantics of a *PROP* predicate such as *PROP*<sub>31</sub> becomes larger and larger. Furthermore, the nodes are more numerous near the leaves and the branching variables play a less important role there. It appeared that one could restrict more the variables to be examined by unit propagation near the leaves without important loss on the search tree size so as to obtain some gain in terms of real run time.

POSIT's UP heuristic (called BCP-based heuristic) [Freeman, 1995] realizes this idea: under the level 9 of a search tree, at most 10 variables are examined by unit propagation.

*Sat*<sub>3141</sub> uses *PROP*<sub>31</sub> from the top of a search tree, but under the depth empirically fixed to  $n * 4/70$ , it uses *PROP*<sub>41</sub>, where *n* is the number of variables in the initial input 3-SAT problem. Note that if  $n \geq 160$ ,  $n * 4/70 \geq 9$ , so *Sat*<sub>3141</sub> generally strengthens the restriction later than POSIT.

From Figures 4 and 5 *Sat*<sub>3141</sub> is not better than *Sat*<sub>31</sub>, although it makes many fewer unit propagations to examine variables (see Figure 6), suggesting that the search tree depth is rather irrelevant to the restriction of UP heuristics.

#### 4.4 Dynamic restriction by number of variables to be examined

The relatively poor performance of *Sat*<sub>42</sub> seems due to the small number of variables examined at each node (see Figure 6), though these variables have many binary occurrences. A careful analysis shows that even *Sat*<sub>31</sub>, the best one up to now, examines few or no variables at some nodes, especially near the root where there are few binary clauses, although these nodes are more determinant for the final search tree size. *PROP*<sub>*T*</sub> is then introduced to ensure that at least *T* variables are examined at each node, *T* being empirically fixed to 10. Near the root, all free variables are examined to exploit the full power of UP heuristic. As soon as the number of variables occurring both negatively and positively in binary clauses and having at least 4 (3) binary occurrences is larger than *T*, only these variables are examined to select the next branching variable.

## 5 Related Work

C-SAT [Dubois et al., 1993] examines some variables by unit propagations (called local processing) near the bottom of a search tree to rapidly detect failed literals there. Pretolani also uses a similar approach (called pruning method) based on hypergraphs in H2R [Pretolani, 1993]. But the local processing and the pruning method as are respectively presented in [Dubois et al., 1993] and [Pretolani, 1993] do not contribute to the heuristic to select the next branching variable. We find the first effective exploitation of UP heuristic in POSIT [Freeman, 1995]

and Tableau [Crawford and Auton, 1996] which use a similar idea as in C-SAT to determine the variables to be examined at a node by unit propagation:  $x$  is to be examined iff  $x$  is among the  $k$  most weighted variables by a Mom's heuristic.

The main difference of Satz with Tableau and POSIT is that Satz does not specify an upper bound  $k$  of the number of variables to be examined at a node by unit propagation. Instead, Satz specifies a lower bound. In fact, Satz examines many more variables by an optimal combination of unit propagation and Mom's heuristics.

Given the depth of a node, Table 1 illustrates the average number of variables examined ( $\#examined\_vars$ ) at the node by Satz, with the depth of the root being 0. In order to compare with C-SAT, Tableau and POSIT we also give the theoretical value of  $k_c$  (for C-SAT),  $k_T$  (for Tableau) and  $k_P$  (for POSIT) at the node, respectively according to the definitions of  $k$  in [Dubois et al., 1993; Crawford and Auton, 1996; Freeman, 1995].

| depth | $\#free\_vars$ | $\#examined\_vars$ | $k_c$ | $k_T$ | $k_P$   |
|-------|----------------|--------------------|-------|-------|---------|
| 1     | 298.24         | 298.24             | 0     | 263   | 265     |
| 2     | 296.52         | 296.52             | 0     | 227   | 230     |
| 3     | 294.92         | 293.89             | 0     | 193   | 198     |
| 4     | 292.44         | 292.21             | 0     | 141   | 149     |
| 5     | 288.60         | 282.04             | 0     | 61    | 72      |
| 6     | 285.36         | 262.14             | 0     | 0     | 10 or 3 |
| 7     | 281.68         | 192.82             | 0     | 0     | 10 or 3 |
| 8     | 277.54         | 125.13             | 0     | 0     | 10 or 3 |
| 9     | 273.17         | 71.51              | 0     | 0     | 10 or 3 |
| 10    | 268.76         | 40.65              | 0     | 0     | 10 or 3 |
| 11    | 264.55         | 26.81              | 0     | 0     | 10 or 3 |
| 12    | 260.53         | 21.55              | 0     | 0     | 10 or 3 |
| 13    | 256.79         | 19.80              | 0     | 0     | 10 or 3 |
| 14    | 253.28         | 19.24              | 0     | 0     | 10 or 3 |
| 15    | 249.96         | 19.16              | 0     | 0     | 10 or 3 |
| 16    | 246.77         | 19.28              | 0     | 0     | 10 or 3 |
| 17    | 243.68         | 19.57              | 0     | 0     | 10 or 3 |
| 18    | 240.68         | 19.97              | 0     | 0     | 10 or 3 |
| 19    | 237.73         | 20.46              | 0     | 0     | 10 or 3 |
| 20    | 234.82         | 20.97              | 0     | 0     | 10 or 3 |

Table 1: Average number of variables examined in Satz at a node in a given depth when solving a hard random 3-SAT problem of 300 variables and 1275 clauses (500 problems are solved) compared with theoretical value of  $k$  in C-SAT, Tableau and POSIT

It is clear that Satz examines many more variables at each node than any of C-SAT, Tableau or POSIT. Near the root, Satz examines all free variables. Elsewhere Satz examines a sufficient number ( $T$ ) of variables.

We compare C-SAT, Tableau, POSIT and Satz on a large sample of hard random 3-SAT problems on a SUN Sparc 20 workstation with a 125 MHz CPU. The 3-SAT problems are generated from 3 sets of  $n$  variables and  $m$  clauses at the ratio  $m/n = 4.25$ ,  $n$  stepping from 300 variables to 400 variables by 50.

We use an executable of C-SAT dated July 1996. The version of Tableau used here is called Stab and is the same used for the experimentation presented in [Crawford and Auton, 1996]. POSIT is compiled using the pro-

vided make command on the SUN Sparc 20 workstation from the sources named `posit-1.0.tar.gz`<sup>1</sup>. Table 2 shows the performances of the 4 DPL procedures on problems of 300, 350, and 400 variables, where time standing for the real mean run time is reported by the unix command `/usr/bin/time` and `t_size` standing for search tree size (number of nodes) is reported (or computed from number of branches reported) by the DPL procedures.

| System  | 300 vars<br>300 problems |        | 350 vars<br>250 problems |        | 400 vars<br>100 problems |         |
|---------|--------------------------|--------|--------------------------|--------|--------------------------|---------|
|         | time                     | t_size | time                     | t_size | time                     | t_size  |
| C-SAT   | 77                       | 49667  | 512                      | 278303 | 3818                     | 1624669 |
| Tableau | 79                       | 43041  | 558                      | 253366 | 4544                     | 1524551 |
| POSIT   | 57                       | 61797  | 474                      | 400588 | 3692                     | 2761611 |
| Satz    | 34                       | 82780  | 203                      | 174337 | 1207                     | 916569  |

Table 2: Mean run time (in second) and mean search tree size of C-SAT, Tableau, POSIT and Satz on ratio  $m/n=4.25$

Table 2 shows that Satz is faster than the above cited versions of C-SAT, Tableau and POSIT, Satz's search tree size is the smallest, and Satz's run time and search tree size grow more slowly. Table 3 shows the gain of Satz compared with the cited version of C-SAT, Tableau and POSIT at the ratio  $m/n=4.25$ . Each item is computed from Table 2 using the following equation:

$$gain = (value(system)/value(Satz) - 1) * 100\%$$

where value is real mean run time or real mean search tree size and system is C-SAT, Tableau or POSIT. From Table 3, it is clear that the gain of Satz grows with the size of the input formula.

| System  | 300 vars<br>300 problems |        | 350 vars<br>250 problems |        | 400 vars<br>100 problems |        |
|---------|--------------------------|--------|--------------------------|--------|--------------------------|--------|
|         | time                     | t_size | time                     | t_size | time                     | t_size |
| C-SAT   | 126%                     | 51%    | 152%                     | 58%    | 216%                     | 77%    |
| Tableau | 132%                     | 31%    | 175%                     | 46%    | 276%                     | 66%    |
| POSIT   | 68%                      | 89%    | 133%                     | 130%   | 198%                     | 200%   |

Table 3: The gain of Satz vs. C-SAT, Tableau and POSIT in terms of run time and search tree size on the ratio  $m/n=4.25$  computed from Table 2

The central strategy of Satz is to try to reach an empty clause as early as possible. Further along the line, we make two relatively small resolvents-driven improvements in Satz. The first improvement is the preprocessing of the input formula by adding some resolvents of length  $\leq 3$ . The second improvement consists in refining yet more the heuristic  $H$  in the nodes where all free variables are examined by unit propagation. Refer to Figure 2, when  $PROP_z$  is equal to  $PROP_0$  we define  $w(x)$  as the number of resolvents the newly produced binary clauses would result in  $F'$  by a single step of resolution.  $w(\bar{x})$  is similarly defined.

<sup>1</sup> publicly available via anonymous ftp to <ftp.cis.upenn.edu> in `pub/freeman/` directory

*Satz* improved in this way solves many real-world or structured SAT problems where previous heuristics were not successful. For example, Table 4 shows the performance of the 4 DPL procedures on the well-known Beijing challenging problems<sup>2</sup>, where a problem that can not be solved in less than 2 hours is marked by "> 7200" and the version of Tableau is called *ntab*<sup>3</sup>. It is clear that *Satz* is much more efficient and solves many more problems in less than two hours.

| Problem          | Satz   | C-SAT  | Posit  | ntab   |
|------------------|--------|--------|--------|--------|
| 2bitadd_10       | > 7200 | > 7200 | > 7200 | > 7200 |
| 2bitadd_11       | 201    | > 7200 | 0.3    | > 7200 |
| 2bitadd_12       | 0.4    | 6379   | 0.05   | > 7200 |
| 2bitcomp_5       | 0.03   | 0.1    | 0.01   | 0.4    |
| 2bitmax_6        | 0.07   | 3.7    | 0.01   | 1.6    |
| 3bitadd_31       | > 7200 | > 7200 | > 7200 | > 7200 |
| 3bitadd_32       | 4512   | > 7200 | > 7200 | > 7200 |
| 3blocks          | 2.0    | 4.3    | 1.6    | 1468   |
| 4blocks          | 8.2    | 118    | 49     | > 7200 |
| 4blocks          | 1542   | > 7200 | > 7200 | > 7200 |
| e0ddr2-10-by-5-1 | 215    | > 7200 | > 7200 | > 7200 |
| e0ddr2-10-by-5-4 | 332    | > 7200 | 3508   | 236    |
| enddr2-10-by-5-1 | > 7200 | > 7200 | > 7200 | > 7200 |
| enddr2-10-by-5-8 | 229    | > 7200 | > 7200 | 92     |
| ewddr2-10-by-5-1 | 339    | > 7200 | 283    | > 7200 |
| ewddr2-10-by-5-8 | 279    | > 7200 | > 7200 | 119    |

Table 4: Run time (in sec.) of Beijing challenging problems

## 6 Conclusion

We found that UP heuristic is substantially better than Mom's one even in its pure form realized by *PROPo* where all free variables are examined at all nodes. In its restricted forms based on combinations of unit propagation and Mom's heuristics, the more variables are examined, the smaller the search tree is, confirming the advantages of UP heuristic, but too many unit propagations slow the execution. The combinations realized by *PROP<sub>41</sub>* and *PROP<sub>31</sub>* represent good compromises.

A dynamic restriction such as *PROP<sub>3141</sub>* which strengthens the restriction under a fixed depth of a search tree fails to work better than the static restriction *PROP<sub>31</sub>*. We design the dynamic restriction along another line: *PROP<sub>z</sub>* ensures that at least *T* candidates are examined by unit propagation at every node of a search tree by successively using *PROP<sub>41</sub>*, *PROP<sub>31</sub>* and *PROPo*, giving the very efficient and very simple DPL procedure called *Satz*.

*Satz* is favorably compared with several current state-of-the-art DPL implementations (C-SAT, Tableau and POSIT) on a large sample of hard random 3-SAT problems and the recent Beijing SAT benchmarks. The good performance of *Satz* on the structured or real-world SAT problems shows that UP heuristic can tackle new problems or problem domains where Mom's heuristics were

not successful and enhances the belief that if a DPL procedure is efficient for random SAT problems, it should be also efficient for a lot of structured ones.

## Acknowledgments

We thank Olivier Dubois, James M. Crawford and Jon W. Freeman for kindly providing us their DPL procedures and anonymous referees for their comments which helped improve this paper.

## References

- [Chvatal and Szemerédi, 1988] V. Chvatal and E. Szemerédi. Many Hard Examples for Resolution. *Journal of ACM*, 35(4):759-768, October 1988.
- [Cook, 1971] S. A. Cook. The Complexity of Theorem Proving Procedures. In *3rd ACM Symp. on Theory of Computing*, pages 151-158, Ohio, 1971.
- [Crawford and Auton, 1996] J. M. Crawford and L. D. Auton. Experimental Results on the Crossover Point in Random 3-SAT. *Artificial Intelligence*, 81, 1996.
- [Davis et al, 1962] M. Davis, G. Logemann, and D. Loveland. *A machine program for theorem proving*. *Communication of ACM*, 5(7):394-397, July 1962.
- [Dubois et al, 1993] Olivier Dubois, P. Andre, Y. Boufkhad and Jacques Carlier. *SAT versus UNSAT Second DIM ACS Challenge: Cliques, Coloring and Satisfiability*, Rutgers University, NJ, 1993.
- [Freeman, 1995] Jon W. Freeman. Improvements to Propositional Satisfiability Search Algorithms. Ph.D. thesis, Department of computer and Information science, Univ. of Pennsylvania, Philadelphia, PA, 1995.
- [Hooker and Vinay, 1995] J. N. Hooker and V. Vinay. Branching Rules for Satisfiability. *Journal of Automated Reasoning*, 15:359-383, 1995.
- [Jeroslow and Wang, 1990] R. Jeroslow and J. Wang. Solving Propositional Satisfiability Problems. *Annals of Mathematics and AI*, 1:167-187, 1990.
- [Li, 1996] ChuMin LI. Exploiting Yet More the Power of Unit Clause Propagation to Solve 3-SAT Problem. In *ECAI'96 Workshop on Advances in Propositional Deduction*, pages 11-16, Budapest, Hungary, 1996.
- [Mitchell et al, 1992] D. Mitchell, B. Selman, H. Levesque. Hard and Easy Distributions of SAT Problems. In *AAAI'92*, pages 459-465, San Jose, CA, 1992.
- [Pretolani, 1993] Daniele Pretolani. Satisfiability and hypergraphs. Ph.D. thesis, Dipartimento di Informatica, Università di Pisa, 1993.

<sup>2</sup>available from <http://www.cirl.uoregon.edu/crawford/beijing>

<sup>3</sup>available from <http://www.cirl.uoregon.edu/crawford/>