



## Sudoku Players' Forums

[FAQ](#) [Search](#) [Memberlist](#) [Usergroups](#)

[Profile](#) [You have no new messages](#) [Log out \[ denis\\_berthier \]](#)

### THE REAL DISTRIBUTION OF MINIMAL PUZZLES

Goto page [Previous](#) [1](#), [2](#), [3](#) ... , [25](#), [26](#), [27](#) [Next](#)



[Sudoku Players' Forums Forum Index -> General/puzzle](#)

[View previous topic](#) :: [View next topic](#)

#### Author

#### Message

**denis\_berthier**

Posted: Wed Sep 30, 2009 5:37 pm Post subject:



Joined: 19 Jun 2007  
Posts: 871  
Location: Paris, France

#### Allan Barker wrote:

#### denis\_berthier wrote:

The statistical results obtained with the controlled-bias generator call for numerous tests:  
- source of complete grids (you and Mike have already done some test with your own sources; gsf's would be another one; I'd like Allan to enter the game because he has a very different generator of complete grids)

If you would like, I could generate 1,000,000 grids and put them on my website, or I could paste the code after a bit of clean up.

As the controlled-bias generator consumes more than 200,000 complete grids per puzzle generated, 1,000,000 grids is far from enough. I fear we need the code, or at least an executable that can output a stream of complete grids.

#### Allan Barker wrote:

It might also be worthwhile to see if they are any good, maybe **Red Ed** could run his bias tester on them?

Didn't he already do it for the complete grids of your rabrnd collection of 1,000,000 puzzles?

[Back to top](#)



**denis\_berthier**

Posted: Wed Sep 30, 2009 6:53 pm Post subject:



Joined: 19 Jun 2007  
Posts: 871  
Location: Paris, France

#### eleven wrote:

With the pre-check, if a 4-cell-unavoidable set has no given (did about 52% of the solves) its about 4.9 times faster than the old version including grid generation.

Is this 1/4.9 ratio to be multiplied by the 1/6 ratio obtained with the first 46 no-test eliminations?

**eleven wrote:**

If no grids file is given, this version prefills the 9 diagonal cells according to Red Ed

I remember vaguely the idea but I can't find the reference for this pre-filling. Doesn't it introduce a bias?

Is it easy to change what is provided: a stream of grids instead of a grids file?

[Back to top](#)



**gsf**

▣ Posted: Wed Sep 30, 2009 7:32 pm Post subject: catalog of all essentially different grids



Joined: 22 Sep 2005  
Posts: 3866  
Location: NJ USA

it turns out these "free" hosting sites are a pain for all but paying customers (no big surprise)  
the worst part is babysitting uploads/downloads *and* the relatively short expirations (30 days) for all that effort

so here's plan B  
this shell script generates the data using my solver  
put this in a file and run "ksh the\_file"

**Code:**

```
typeset -Z3 band
for ((band=1; band<300; band++))
do
    if [[ ! -f $$band.sudz ]]
    then
        sudoku -gb$band -f%#ec > $band.tmp
        mv $band.tmp $band.sudz
        print -u2 $band.sudz
    fi
done
```

this will generate 001.sudz .. 299.sudz, printing each band file name on stderr as it completes

I already posted 300-416.sudz, so with that you will have the complete catalog

you can re-start the script and it will pick up where it left off  
so you could run it overnight when you are sleeping (or texting)

this will probably take ~2 weeks on most laptops

one more note on the sudz compression  
it uses some domain knowledge to increase compression with a space/time tradeoff  
a simple (and fast) singles solver is used to identify implicit clues after the prefix (known clues in earlier rows)  
and only explicit clues are compressed  
the implicit clues are recovered during decompression

Last edited by gsf on Thu Oct 01, 2009 6:19 am; edited 1 time in total

[Back to top](#)

[profile](#) [pm](#) [www](#)

**Red Ed**

Posted: Wed Sep 30, 2009 9:19 pm Post subject:

[quote](#)

Joined: 06 Jun 2005  
Posts: 747

**denis\_berthier wrote:**

**eleven wrote:**

If no grids file is given, this version prefills the 9 diagonal cells according to Red Ed

I remember vaguely the idea but I can't find the reference for this pre-filling. Doesn't it introduce a bias?

Yes, it introduces bias. The bias in the statistics computed by my tester is less than in the original *suexg*. [<link>](#)

[Back to top](#)

[profile](#) [pm](#)

**denis\_berthier**

Posted: Thu Oct 01, 2009 4:04 am Post subject:

[quote](#) [edit](#)

Joined: 19 Jun 2007  
Posts: 871  
Location: Paris, France

**Red Ed wrote:**

**denis\_berthier wrote:**

**eleven wrote:**

If no grids file is given, this version prefills the 9 diagonal cells according to Red Ed

I remember vaguely the idea but I can't find the reference for this pre-filling. Doesn't it introduce a bias?

Yes, it introduces bias. The bias in the statistics computed by my tester is less than in the original *suexg*. [<link>](#)

Just to be sure: the diagonal is filled randomly?

[My problem comes in part from things like:  $k=(MWC > 9) \& 15$ . I don't know how to interpret the  $\& 15$ . ]

Last edited by denis\_berthier on Thu Oct 01, 2009 5:20 am; edited 1 time in total

[Back to top](#)

[profile](#) [pm](#) [www](#)

**denis\_berthier**

Posted: Thu Oct 01, 2009 4:26 am Post subject: Re: catalog of all essentially different grids

[quote](#) [edit](#)

Joined: 19 Jun 2007  
Posts: 871  
Location: Paris, France

**gsf**, thanks for all.

10/01/09, 6 am (Paris time): I've just launched the script. (I had to delete the double quotes around the redirection sign.)

We'll see how long the whole thing takes.

I got the first 001.sudz (1 Mb) and checked that I could decompress it correctly.

Decompression takes only ~15 s. to produce 1007170 puzzles (occupying 82.6 Mb).

[Back to top](#)



**denis\_berthier**

Posted: Thu Oct 01, 2009 6:13 am Post subject:



Joined: 19 Jun 2007  
Posts: 871  
Location: Paris, France

Thinking again of the (very small) bias due to automorphisms in gsf's complete grids list, there's a conceptually easy way of eliminating it. As gsf's decompression algorithm can output a stream of (puzzle, automorphism-count) pairs, the input to the controlled bias generator could be modified to take this couple as input instead of just the puzzle. It would re-use each puzzle the number of times defined by automorphism-count [Edit after Red Ed's remark: i.e. use it with probability 1/automorphism-count]

More generally, I think that suexg-cb could advantageously be split into 2 separate parts: generation, deletion, that could be combined via Unix piping. That would make testing different combinations easier, especially now that we have several sources of complete grids and several efficiency improvements for the deletion part. (It doesn't matter is the 'solve' function has to appear in the two parts.)

The generation parts could be written in a uniform way as outputting a stream of complete grids or a stream of (complete-grid, number) pairs. (The entries - such as: number of desired grids, RNG seed - could depend on the particular type of generator, maybe even the RNG could be chosen as a parameter, so as to minimise the number of different programs).

The deletion part could be written in a uniform way as taking an input stream of either grids or (grid, number) pairs, with number defaulting to 1 [Edit after Red Ed's remark: and grid is used with probability 1/number], and outputting a stream of controlled-bias minimal puzzles - or (minimal-puzzle, number-of-grids-used) pairs. (Notice that the format of the input stream elements is the same for the whole stream: always a puzzle or always a pair; so that no test would have to be repeated).

I don't know if "conceptually simple" entails easy to implement (e.g. can an input stream have a non predefined number of arguments? or should this number be passed as an additional argument to the deletion program?) That's obviously a question for eleven.

Last edited by denis\_berthier on Thu Oct 01, 2009 7:11 am; edited 2 times in total

[Back to top](#)



**gsf**

Posted: Thu Oct 01, 2009 6:29 am Post subject: Re: catalog of all essentially different grids



Joined: 22 Sep 2005  
Posts: 3866  
Location: NJ USA

**denis\_berthier wrote:**

**gsf**, thanks for all.  
10/01/09, 6 am (Paris time): I've just launched the script. (I had to

delete the double quotes around the redirection sign.)

rats -- that was a remnant from testing the script  
001.sudz is one of the smaller bands  
the largest band is 006.sudz @ 79,749,100 bytes

[Back to top](#)

[profile](#) [pm](#) [www](#)

**Red Ed**

Posted: Thu Oct 01, 2009 6:51 am Post subject:

[quote](#)

Joined: 06 Jun 2005  
Posts: 747

**denis\_berthier wrote:**

Thinking again of the (very small) bias due to automorphisms in gsf's complete grids list, there's a conceptually easy way of eliminating it. As gsf's decompression algorithm can output a stream of (puzzle, automorphism-count) pairs, the input to the controlled bias generator could be modified to take this couple as input instead of just the puzzle. It would re-use each puzzle the number of times defined by automorphism-count.

"Defined by" - yes; "equal to" - no. 😊 The expected number of times that any solution grid is reused should be proportional to  $1/\#auts$ , since each canonicalised gsf grid gives rise to  $3359232/\#auts$  non-canonicalised grids (the sum of which counts gives  $\sim 6.67e21$ ).

[Back to top](#)

[profile](#) [pm](#)

**Red Ed**

Posted: Thu Oct 01, 2009 6:53 am Post subject:

[quote](#)

Joined: 06 Jun 2005  
Posts: 747

btw, can anyone remember if we ever found a minimal puzzle that was automorphic?

[Back to top](#)

[profile](#) [pm](#)

**gsf**

Posted: Thu Oct 01, 2009 7:02 am Post subject:

[quote](#)

Joined: 22 Sep 2005  
Posts: 3866  
Location: NJ USA

**Red Ed wrote:**

btw, can anyone remember if we ever found a minimal puzzle that was automorphic?

I just checked the patterns game results and there were 32  
here's one with 4 automorphisms  
(I would appreciate a double check on that)

**Code:**

```
. . . 1 . 2 . 3 .
3 . . . . 4 . .
. 5 1 . . 4 6 . .
7 . 5 2 . 8 . . 6
. . . . . . . .
4 . . 7 . 9 1 . 8
. . 4 6 . . 5 1 .
. . 6 . . . . 3
. 3 . 9 . 5 . . .
```

[Back to top](#)**denis\_berthier**

Posted: Thu Oct 01, 2009 7:04 am Post subject:



Joined: 19 Jun 2007  
 Posts: 871  
 Location: Paris, France

**Red Ed wrote:****denis\_berthier wrote:**

Thinking again of the (very small) bias due to automorphisms in gsf's complete grids list, there's a conceptually easy way of eliminating it.

As gsf's decompression algorithm can output a stream of (puzzle, automorphism-count) pairs, the input to the controlled bias generator could be modified to take this couple as input instead of just the puzzle. It would re-use each puzzle the number of times defined by automorphism-count.

"Defined by" - yes; "equal to" - no. 😊 The expected number of times that any solution grid is reused should be proportional to  $1/\#auts$ , since each canonicalised gsf grid gives rise to  $3359232/\#auts$  non-canonicalised grids (the sum of which counts gives  $\sim 6.67e21$ ).

Oops, yes, I'll correct my previous post to eliminate any ambiguity.

[Back to top](#)**Red Ed**

Posted: Thu Oct 01, 2009 7:28 am Post subject:



Joined: 06 Jun 2005  
 Posts: 747

**gsf wrote:****Red Ed wrote:**

btw, can anyone remember if we ever found a minimal puzzle that was automorphic?

I just checked the patterns game results and there were 32 here's one with 4 automorphisms  
 (I would appreciate a double check on that)

Yep, that works. Nice one.

[Back to top](#)**m\_b\_metcalf**

Posted: Thu Oct 01, 2009 8:02 am Post subject:



Joined: 15 May 2006  
 Posts: 2399  
 Location: Berlin

**Red Ed wrote:**

Yep, that works. Nice one.

From Mauricio, 10.2/1.2/1.2.

Regards,

Mike Metcalf

[Back to top](#)

**eleven**

Posted: Thu Oct 01, 2009 11:15 am Post subject:

Joined: 10 Feb 2008  
Posts: 503

Denis,

if you want to pipe a stream of grids to the cb generator, you can try this:  
Replace**Code:**

```
if(argc>3)if((file=fopen(argv[3],"rb"))==NULL)
                {printf("\ncan't find file
%s\n",argv[3]);return(1);}
```

by

**Code:**

```
if(argc>3){
    if (strcmp(argv[3], "-") == 0)
        file = stdin;
    else if((file=fopen(argv[3],"rb"))==NULL)
        {printf("\ncan't find file
%s\n",argv[3]);return(1);}}
```

Then it reads the grids from stdin, if you enter '-' as file name, e.g.  
./grid\_generator | ./cb\_generator 0 1 - > puzzles.dat[Back to top](#)Display posts from previous:   **Sudoku Players'****Forums Forum**  
Index ->  
**General/puzzle**All times are GMT  
Goto page [Previous](#) [1](#), [2](#), [3](#) ... , [25](#), [26](#), [27](#) [Next](#)**Page 26 of 27**[Stop watching this topic](#)Jump to:  

You **can** post new topics in this forum  
 You **can** reply to topics in this forum  
 You **can** edit your posts in this forum  
 You **can** delete your posts in this forum  
 You **can** vote in polls in this forum

Powered by phpBB © 2001, 2005 phpBB Group